



UFPB

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS AGRÁRIAS**

**DEPARTAMENTO DE FITOTECNIA E CIÊNCIAS AMBIENTAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**



T E S E

**MARCADORES ESPECTRAIS NO VISÍVEL E INFRAVERMELHO
PRÓXIMO PARA IDENTIFICAÇÃO DE SEMENTES EM PROGÊNIES
DE *Ricinus communis* L.**

DANIEL DA SILVA FERREIRA

**Areia, PB
Junho de 2015**



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE FITOTECNIA E CIÊNCIAS AMBIENTAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA**



**MARCADORES ESPECTRAIS NO VISÍVEL E INFRAVERMELHO
PRÓXIMO PARA IDENTIFICAÇÃO DE SEMENTES EM PROGÊNIES
DE *Ricinus communis* L.**

DANIEL DA SILVA FERREIRA

Sob o comitê de Orientação dos Professores

Prof^a. Dr^a. Riselane de Lucena Alcântara Bruno

e

Prof. Dr. Everaldo Paulo de Medeiros

Tese submetida como requisito para
obtenção do grau de **Doutor em
Agronomia**, no Programa de Pós-
Graduação em Agronomia

Areia, PB
Junho de 2015

Catálogo na publicação
Seção de Catalogação e Classificação

F383m FERREIRA, DANIEL DA SILVA.

MARCADORES ESPECTRAIS NO VISÍVEL E INFRA-VERMELHO
PRÓXIMO PARA IDENTIFICAÇÃO DE SEMENTES EM PROGENIES DE
RICINUS COMMUNIS L. / Daniel da Silva Ferreira. -
AREIA, 2015.
134 f.

Orientação: RISELANE DE LUCENA ALCÂNTARA BRUNO.
Coorientação: EVERALDO PAULO DE MEDEIROS.
Tese (Doutorado) - UFPB/CCA.

1. RICINUS COMMUNIS L. 2. CLASSIFICAÇÃO DE SEMENTES. 3.
ESPECTROSCOPIA NO INFRA-VERMELHO PRÓXIMO-NIR. 4.
ESPECTROSCOPIA NO VISÍVEL. I. BRUNO, RISELANE DE LUCENA
ALCÂNTARA. II. MEDEIROS, EVERALDO PAULO DE. III. Título.

UFPB/CCA-AREIA



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE FITOTECNIA E CIÊNCIAS AMBIENTAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

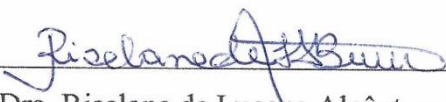



CERTIFICADO DE APROVAÇÃO

TÍTULO: Marcadores espectrais no visível e infravermelho próximo para identificação de sementes em progênies de *Ricinus communis* L.

AUTOR: DANIEL DA SILVA FERREIRA

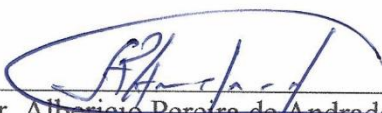
Aprovado como parte das exigências para obtenção do título de DOUTOR em AGRONOMIA (Ciência e Tecnologia de Sementes, Bioquímica e Fisiologia Pós-Colheita) pela comissão Examinadora:

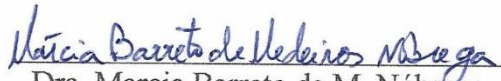

Prof. Dra. Riselane de Lucena Alcântara Bruno

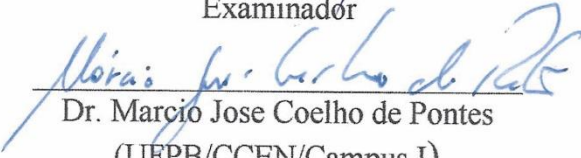

Prof. Dr. Everaldo Paulo de Medeiros
Comitê de Orientação

Data da realização: 26 de junho de 2015

Dra. Riselane de Lucena Alcântara Bruno
(UFPB/CCA/Campus II)
Orientadora


Dr. Alberício Pereira de Andrade
(UFPB/CCA/Campus II)
Examinador


Dra. Marcia Barreto de M. Nóbrega
(EMBRAPA/Embrapa Algodão)
Examinadora


Dr. Marcio Jose Coelho de Pontes
(UFPB/CCEN/Campus I)
Examinador

Presidente da Comissão Examinadora
Prof. Dra. Riselane de Lucena Alcântara Bruno
Orientadora

DEDICATÓRIA

*Dedico mais essa vitória a **Deus** que em sua infinita bondade, está sempre ao meu lado, me dando muito mais do que mereço e atendendo aos pedidos do meu coração. Obrigado meu Deus!*

*Aos meus pais **Otoniel** (in memoriam) e **Maria Luiza**, certamente os grandes responsáveis pela minha vida, pela pessoa que me tornei e por tudo que conquistei. Os ensinamentos que me foram dados, e as referências transmitidas, servem de alíquotas para a construção do meu ser. Amo vocês!*

*A minha esposa **Tanísia** que tanto me apoiou e soube identificar os momentos onde a força começava a dar espaço a fragilidade. Seu carinho e incentivo certamente foram essenciais para fortalecer minha vontade de vencer as imensas dificuldades desta trajetória. Juntos nos completamos!!!*

*Aos meus filhos **Marcus Paulo**, **Rauê Lucas** e **Davi Luiz** que souberam mesmo em sua inocência entender a ausência do pai e nos momentos mais difíceis sempre estavam com sorriso aberto e gestos de carinho gratuito a oferecer, contribuindo para retirada de todas as tristezas e angústias do meu coração. Razão do meu viver!!!*

*A minha irmã **Ana** e meus sobrinhos **Edjhon** e **Edjalma** pela franqueza da nossa relação, e pela amizade sempre presente. Nossas diferenças nos fortalecem!!!*

*A Sr. **José Rodrigues**, um exemplo de força e incentivo, onde a palavra “talvez” ou mesmo “não consigo” praticamente inexistem. A conquista depende da nossa vontade de vencer!!*

*Aos meus **familiars** e **amigos** por estarem presentes e pertencerem a escrita de mais estas páginas da minha história. A amizade é o amor que nunca morre!!*

AGRADECIMENTO

“Gratidão é uma forma singular de reconhecimento, e o reconhecimento é uma forma sincera de gratidão”.

Alan Vaszattz

A Deus por transformar minhas fraquezas em fortaleza, minha desistência em perseverança, meus medos em coragem, minhas dúvidas em certezas e me fazer acreditar sempre em seu amor, justiça e bondade.

A minha Família, minha gratidão por me incentivarem, me darem força, apoio, carinho, atenção, compreensão em todas as horas e acreditarem em mim, frente as dificuldades encontradas.

A minha orientadora Profa. Dra Ríselane de Lucena Ancântara Bruno, minha gratidão por tantos conhecimentos transmitidos, pelas orientações técnicas e científicas repassadas e ainda as orientações pessoais ofertadas. Por ter me recebido como aluno, pelo carinho, paciência, compreensão, amizade e pelo exemplo profissional e pessoal que merece reconhecimento de todos a sua volta. Minha eterna gratidão!

A meu Co-orientador Dr. Everaldo Paulo de Medeiros, pelo exemplo profissional e de ser humano, pela atenção, humildade, profissionalismo, responsabilidade, respeito, ética e apoio incessante na condução dos trabalhos, certamente um aprendizado para toda a vida. Meu muito obrigado!

A Universidade Federal da Paraíba, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia (PPGA) pela oportunidade concedida para o engrandecimento para minha formação pessoal e profissional em especial a Profa. Dra Edna Ursulino Alves, a Profa. Dra Luciana Cordeiro e a Eliane Araújo, pela atenção e a pré-disposição em ajudar nas horas mais difíceis.

A Empresa Brasileira de Pesquisa Agropecuária-EMBRAPA em especial a Embrapa Algodão e a Embrapa Produtos e Mercado, pelo apoio nas diversas etapas de desenvolvimento dos trabalhos.

A Dra. Marcia Barreto de Medeiros Nobrega, Dra Maira Milani, ao Prof. Dr Mauro Nobrega da Costa, ao Pesquisador Francisco Pereira de Andrade, a técnica Maria do Carmo de Lucena Leite e a assistente Karina Rione Silva Gomes e Sousa Araujo, por estarem sempre de portas abertas para ajudar, esclarecer as dúvidas, me incentivar e ainda na cessão das sementes utilizadas nas avaliações, meu muito obrigado pelo profissionalismo de sempre.

A equipe do LATCCQ, em especial aos técnicos Joabson Borges de Araujo, Edijane Valeria Araujo dos Anjos e a bolsista Iranilma Maciel, que me acolheram, contribuíram definitivamente nas atividades, me incentivaram e torceram por mim.

A equipe do SPM-Escritório de Campina Grande, que me deram total e irrestrito apoio para que fosse possível a realização de mais este sonho, minha gratidão a Washington Luiz, Joana D'arc, Rita de Cassia, José Marcelo, Ernani Vidal, Danilo Alves e a Sra. Edivam Mousinho.

Aos meus irmãos do coração Amanda Kelly, Aderdilândia Iane e Givanildo Zildo, pela amizade, carinho, força e alegria mesmo no sol a pino desta jornada e por compartilharmos varias de nossas alegrias, dificuldades e angustias. Vocês são verdadeiros amigos!

A Demetrius Silva e Francisco Gambarra (Chicão) pela amizade e sempre disponibilidade de ajudar, “caba vei” estamos juntos, meu muito obrigado!

Ao Prof. Dr. Albericio Pereira de Andrade, a Dra. Marcia Barreto de Medeiros Nobrega e ao Prof. Dr. Marcio Pontes que aceitaram de imediato ao nosso convite para a banca examinadora e certamente só vem a contribuir com mais essa conquista. Muito obrigado!!!

E nesse momento em que se fecha mais um ciclo agradeço a todos os amigos, familiares e colegas que participaram, torceram e vibraram. Com certeza estão em festa com mais essa vitória. Não sou muito bom de despedidas pois me enquadro bem no que diz Caio Fernando Abreu...

“Sabz, eu acho que não sei fechar ciclos, colocar pontos finais. Comigo são sempre virgulas, aspas, reticências... eu vou gostando... eu vou cuidando, eu vou desculpando, eu vou superando, eu vou compreendendo, eu vou relevando, eu vou... e continuo indo, assim, desse jeito, sem virar páginas, sem colocar pontos...”

Obrigado a todos!!!

SUMARIO

SUMARIO.....	viii
LISTA DE FIGURAS.....	x
LISTA DE TABELAS.....	xiv
RESUMO GERAL.....	15
GENERAL ABSTRACT.....	16
1. INTRODUÇÃO GERAL.....	17
2. FUNDAMENTAÇÃO TEORICA.....	19
2.1. A produção de sementes de mamona	19
2.2. Normas para a produção de sementes de mamona	19
2.3. Classificação de sementes de mamona.....	21
2.4. Espectroscopia na região do visível.....	22
2.5. Espectrometria no infravermelho	23
2.5.1. Vibrações moleculares	25
2.6. Espectrometria no infravermelho próximo – NIR	28
2.6.1. A espectroscopia NIR.....	28
2.7. Região infravermelho próximo de ondas curtas-SWNIR.....	32
2.8. Avaliações das bandas espectrais	33
2.9. Principais aplicações do NIR.....	33
2.10. Quimiometria	35
2.10.1. Pré processamento	36
2.10.2. Reconhecimento de padrões não supervisionados	38
2.10.3. Reconhecimento de padrões supervisionados	40
2.11. Uso da espectroscopia NIR em sementes na diferenciação de genótipos	45
3. REFERÊNCIAS BIBLIOGRÁFICAS	47
 CAPITULO I MARCADORES ESPECTRAIS SWNIR EM CULTIVARES DE MAMONA	64
RESUMO	65
ABSTRACT	66
1. INTRODUÇÃO	67
2. MATERIAL E MÉTODOS	71

2.1.	Aquisição das Amostras	71
2.2.	Instrumentação	72
2.3.	Aquisição dos espectros NIR	73
2.4.	Programas Computacionais	74
2.5.	Tratamento Quimiométrico dos Dados.....	74
3.	RESULTADOS E DISCUSSÃO	75
3.1.	Espectros NIR	75
3.2.	Análise Exploratória dos Dados.....	79
3.3.	Reconhecimento de Padrões Supervisionados	84
4.	CONCLUSÕES.....	90
5.	REFERENCIAS BIBLIOGRAFICAS	91
CAPITULO II MARCADORES ESPECTRAIS VIS-NIR EM PROGENIES DE MAMONA.....		97
RESUMO		98
ABSTRACT		99
1.	INTRODUÇÃO	100
2.	MATERIAL E METODOS	103
2.1.	Aquisição das Amostras	103
2.2.	Instrumentação	104
2.3.	Aquisição dos espectros NIR	105
2.4.	Programas Computacionais	105
2.5.	Tratamento Quimiométrico dos Dados.....	105
3.	RESULTADOS E DISCUSSÃO	107
3.1.	Espectros NIR	107
3.2.	Análise Exploratória dos Dados.....	111
3.3.2.1.	Validação cruzada.....	126
3.3.2.2.	Validação por serie de teste.....	128
4.	CONCLUSÕES.....	132
5.	REFERENCIAS BIBLIOGRAFICAS	133

LISTA DE FIGURAS

Figura 1- Oscilações moleculares de deformação.	26
Figura 2- Representação sistemática da energia potencial para (a) oscilação harmônico, (b) oscilador anarmônico. De = distância do equilíbrio (E = mínimo)..	27
Figura 3- Modos de medição utilizados em espectroscopia NIR. 11 (A) transmitância; 11 (B) transfectância e 11 (C) reflectância difusa, através do meio de dispersão.	31
Figura 4- Principais bandas de absorção (Frequência de grupos funcionais) na região NIR e sua localização no espectro eletromagnético.....	33
Figura 5- Plano bidimensional com duas variáveis e representação das duas primeiras componentes principais..	40
Figura 6- Sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.	71
Figura 7- Espectrofotômetro VIS-NIR.....	72
Figura 8- (A) Célula de quartzo; (B) Tampas reflexivas para a célula de quartzo.....	73
Figura 9- Padrão de reflectância.....	73
Figura 10- Grupos estabelecidos (A) BRS Nordestina (I) e BRS Paraguaçu (II) e grupo (B) BRS Energia (I), linhagem CNPAM 2009-7(II) e BRS Gabriela III).	75
Figura 11- Espectros originais NIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7, na faixa de 400-2500nm.	76
Figura 12- Espectros originais SWNIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.	76
Figura 13- Espectros NIR derivativos das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7, na faixa de 400-2500nm.	77
Figura 14- Espectros originais médios SWNIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.	77
Figura 15- Espectros médios NIR de reflectância difusa pré-processados das sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.....	78

Figura 16- Espectros médios SWNIR de reflectância difusa pré-processados das sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.....	79
Figura 17- Gráfico dos escores (PC1 vs PC2) para o conjunto das 250 amostras de sementes de mamona BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.	80
Figura 18- Gráfico de pesos na PC1 para o conjunto das 250 amostras de sementes de mamona BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.	80
Figura 19- Gráfico dos escores (PC1 vs PC2) para o conjunto das 100 amostras de sementes de mamona das cultivares BRS Nordestina e BRS Paraguaçu.....	81
Figura 20- Gráfico de pesos (PC1) para o conjunto das 100 amostras de sementes de mamona BRS Nordestina e BRS Paraguaçu.	82
Figura 21- Gráfico dos escores (PC1 vs PC2) para o conjunto das 150 amostras de sementes de mamona BRS Energia, BRS Gabriela e CNPAM 2009-7.	83
Figura 22- Gráfico de pesos (PC1) para o conjunto das 150 amostras de sementes de mamona BRS Energia, BRS Gabriela e CNPAM 2009-7.	84
Figura 23- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA...87	
Figura 24- Espectro médio das amostras de treinamento com as variáveis selecionadas pelo SPA-LDA.	89
Figura 25- Sementes de mamona de acordo com seus parentais e suas gerações recíprocas.	103
Figura 26- Espectrofotômetro VIS-NIR.....	104
Figura 27- (A) Célula de quartzo; 36 (B) tampas reflexivas para a célula de quartzo.....	104
Figura 28- Padrão de reflectância.....	105
Figura 29- Espectros originais NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500nm	107
Figura 30- Espectros derivativos NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500nm.	108
Figura 31- Espectros médios originais NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e cruzamentos recíprocos obtidos entre 400 e 2500nm.	108
Figura 32- Espectros médios derivativos NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e cruzamentos recíprocos obtidos entre 400 e 2500nm.	109

Figura 33- Espectros originais VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.....	109
Figura 34- Espectros derivativos VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.....	110
Figura 35- Espectros médios originais VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.	110
Figura 36- Espectros médios derivativos VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.	111
Figura 37- Gráfico dos escores em A (PC1 vs PC2) e em B (PC1 vs PC3) para o conjunto de 200 amostras como resultado da aplicação da PCA a todos os espectros pré-processados. ..	112
Figura 38- Gráfico dos pesos (PC1) para o conjunto de 200 amostras como resultado da aplicação da PCA a todos os espectros pré-processados.....	112
Figura 39- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos genitores CNPAM 2012-137e CNPAM 93168.....	113
Figura 40- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos genitores CNPAM 2012-137 e CNPAM 93-168	113
Figura 41- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F1.	114
Figura 42- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F1.	114
Figura 43- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F2.	115
Figura 44- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F2.	115
Figura 45- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F3.	116
Figura 46- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F3.	117
Figura 47- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F4.	118
Figura 48- Gráfico de pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F4.	118

Figura 49- Gráfico dos escores (PC1 vs PC2) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 2012-137e seus cruzamentos F1, F2, F3 e F4.....	119
Figura 50- Gráfico dos pesos (PC1) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 2012-137e seus cruzamentos F1, F2, F3 e F4.....	120
Figura 51- Gráfico dos escores (PC1 vs PC2) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 93-168 e seus cruzamentos F1, F2, F3 e F4.....	121
Figura 52- Gráfico dos pesos (PC1) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 93-168 e seus cruzamentos F1, F2, F3 e F4.....	121
Figura 53- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA por validação cruzada em (a) visualização normal e em (b) com efeito do zoom.....	126
Figura 54- Espectro médio dos genótipos com as variáveis selecionada pelo SPA-LDA. ...	126
Figura 55- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA, por série de teste.....	129
Figura 56- Espectro médio das amostras de treinamento com as variáveis selecionada pelo SPA-LDA.	131

LISTA DE TABELAS

Tabela 1. Parâmetros de Campo utilizados na produção de sementes de mamona.....	20
Tabela 2. Parâmetros de semente utilizados na produção de sementes de mamona.	21
Tabela 3. Frações de absorbância e transmitância específicas e suas correspondentes nas faixas espectrais pertencentes a região do visível.	23
Tabela 4. Limites aproximados das três regiões do infravermelho.	24
Tabela 5. Limites aproximados das duas regiões do infravermelho.	30
Tabela 6. Caracterização morfológica das sementes de mamona.....	72
Tabela 7. Número de amostras dos conjuntos de validação cruzada, selecionadas pelo algoritmo KS, para as classes BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e uma linhagem CNPAM 2009-7.	75
Tabela 8. Resumo de erros de classificação obtido pelos modelos SIMCA utilizando-se o conjunto de amostras de previsão das sementes de mamona nos níveis de significância do Teste – F (1, 5, 10 e 25%).....	85
Tabela 9. Número de amostras dos conjuntos de validação cruzada e teste, selecionadas pelo algoritmo KS, para os progenitores e os cruzamentos recíprocos.	106
Tabela 10. Número de amostras dos conjuntos de validação cruzada e teste, selecionadas pelo algoritmo KS, para os progenitores e os cruzamentos recíprocos.	107
Tabela 11. Erros de classificação ao nível de 1, 5, 10 e 25% de confiança para teste F.	123
Tabela 12. Erros de classificação ao nível de 5% de confiança para teste F.	125
Tabela 13. Resultados do algoritmo SPA-LDA para o conjunto de validação cruzada separado pelo algoritmo KS.	127
Tabela 14. Resultados do algoritmo SPA-LDA para o conjunto de previsão separado pelo algoritmo KS.	129

FERREIRA, D.F. **Estudo de marcadores espectrais no visível e infravermelho próximo para seleção de sementes de mamona (*Ricinus communis* L.)**. 2015. 134 f. Tese (Doutorado em Agronomia), Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba.

RESUMO GERAL

A produção e a disponibilidade de sementes de mamona (*Ricinus communis* L.) com qualidade física, genética e fisiológica, constitui um desafio para a cadeia produtiva, pois, apesar da cultura possuir variabilidade relativas as suas sementes, em muitos casos não é possível a distinção visual de alguns genótipos o que pode promover a perda de qualidade. Com o objetivo de se estabelecer marcadores espectrais de identificação de sementes de mamona baseado em cultivares já estáveis e em gerações envolvendo, diferentes progenitores utilizando a espectroscopia, é que foi desenvolvido o presente estudo. A espectroscopia consiste em um método analítico, onde é estudada a interação das radiações eletromagnéticas com as moléculas, sendo explorada neste estudo a região do visível VIS-NIR, com faixa de absorção aproximada entre 400 a 750nm e a faixa do infravermelho próximo de ondas curtas SWNIR compreendida entre 750 a 1100nm. No primeiro estudo, foram avaliadas quatro cultivares sendo: BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e a linhagem CNPAM 2009-7. Os espectros de reflectância foram registrados na faixa de 750 a 1100nm. Para os modelos de classificação foram utilizadas 50 sementes de cada material, sendo selecionadas para o conjunto de teste 36 sementes de cada. No segundo estudo avaliou-se dois genitores sendo: CNPAM 2012-137 e CNPAM 93168, e suas gerações de recíprocas de F1, F2, F3 e F4. Os espectros de reflectância foram registrados na faixa do visível VIS-NIR de 400 a 750nm. Para os modelos de classificação foram utilizadas 20 sementes de cada material, e para o conjunto de teste 10 sementes de cada. As leituras foram efetuadas com as sementes sempre na mesma posição. Os espectros foram pré-processados utilizando o algoritmo Savitzky-Golay com janela de 11 pontos e primeira derivada para correção da linha de base. Mediante a PCA (Principal Component Analysis) foi constatada que tanto na faixa SWNIR quanto na faixa VIS-NIR em estudo, os resultados permitiram a distinção entre as cultivares e as gerações recíprocas. O modelo SIMCA demonstrou resultados promissores para classificação para os níveis de significância de 1,5 e 10% em ambas as faixas de estudo. O SPA-LDA (Sucessive Projections Algorithm-Linear Discriminant Analysis) foi eficiente classificando corretamente todas as amostras, tanto na região do SWNIR quanto no VIS-NIR. Os métodos em estudo foram comprovadamente promissores quanto a diferenciação fenotípica para sementes de mamoneira e suas gerações recíprocas, de forma rápida, prática e não destrutiva.

Palavras-chave: Produção de sementes, *Ricinus communis* L., Espectroscopia SWNIR, Espectroscopia VIS-NIR

FERREIRA, D.F. **Study spectral labels in the visible and infrared near for selection of seeds castor bean (*Ricinus communis* L.).** 2015. 134f. Thesis (Doctorate in Agronomy Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba.

GENERAL ABSTRACT

The production and availability of seeds of castor bean (*Ricinus comunnis* L.) with physical, genetic and physiological quality, constitutes a challenge to the productive chain, because, despite their own variability concerning seeds, in many cases it is not possible to distinguish visually some genotypes that which can promote the loss of quality. With the aim of establishing spectral markers based on identification of castor seeds already stable and cultivars generations involving different progenitors using spectroscopy, is that this study was developed. Spectroscopy consists of an analytical method, where it is studied the interaction of electromagnetic radiation with the molecules being explored in this study the visible region VIS-NIR, with approximate absorption range between 400 to 750nm and near shortwave infrared band SWNIR between 750 to 1100nm. In the first study, four cultivars were evaluated: BRS Nordeste, BRS Paraguaçu, BRS Energy, BRS Gabriela and the lineage CNPAM 2009-7. The reflectance spectra were recorded in the range of 750 to 1100nm. For Classification models that were used 50 seeds of each material, being selected for the test suite 36 seeds each. The second study evaluated two genitors being: CNPAM 2012-137 and CNPAM 93-168, and their reciprocal generations of F1, F2, F3 and F4. The reflectance spectra were recorded in the visible band VIS-NIR of 400 to 750nm. For classification models were used 20 seeds of each material, and for the test set 10 seeds of each. The readings were made with the seeds always in the same position. The spectra were preprocessed using the algorithm Savitzky-Golay with window of 11 points and first derivative for baseline correction. By PCA (Principal Component Analysis) was established that both the SWNIR band as on VIS-NIR band in study, the results allowed the distinction between the cultivars and reciprocal generations. The SIMCA model demonstrated promising results for classification to the significance levels of 1.5 and 10% in both tracks of study. The SPA-LDA (Successive Projections Algorithm-Linear Discriminant Analysis) was efficient classifying correctly all the samples, both in the SWNIR and in the VIS-NIR band. The methods in study have shown to be promising for the phenotypic differentiation of castor bean seeds and their reciprocal generations in a fast, practical and non-destructive.

Keywords: seed production, *Ricinus comunnis* L., SWNIR spectroscopy, spectroscopy VIS-NIR

1. INTRODUÇÃO GERAL

As sementes constituem um dos mais importantes aportes tecnológicos, para que seja iniciada qualquer atividade agrícola produtiva, é por intermédio destas que serão traçadas todas as estratégias buscando obter os melhores resultados possíveis, e para isto, se faz necessário o uso de sementes com origem conhecida e qualidade garantida.

Em geral, a qualidade das sementes muitas vezes é associada ao aspecto visual e não reproduzem em campo o resultado esperado por ocasião da aquisição, uma vez que, a qualidade envolve diversos atributos que estão intimamente ligados ao sistema de produção das mesmas.

Os atributos genéticos, físicos, fisiológicos e de sanidade compõem a qualidade das sementes e interferem em seu potencial de desempenho. Considerando os prejuízos que a presença de “sementes contaminantes” pode causar, após sua multiplicação, justifica-se a adoção de programa de controle de qualidade envolvendo todas as etapas de produção com o intuito de assegurar que as características selecionadas pelos melhoristas sejam transferidas para o produtor rural (MARCOS FILHO, 1999).

Segundo Carvalho & Nakagawa (1988) a injúria mecânica, juntamente com a mistura varietal, é apontada pelos tecnologistas como um dos mais sérios problemas da produção de sementes. Essa mistura pode ocorrer tanto na produção por parte do obtentor como também nas etapas conduzidas pelos multiplicadores, isso ocorre, por exemplo, por falta de isolamento entre cultivares distintas, limpeza ineficiente dos implementos agrícolas, permanência de sementes em máquinas e instalações para semeadura, colheita, secagem, beneficiamento, transporte e armazenamento (BORÉM & CAIXETA, 2009; RAMOS, 2004).

Assim, considera-se a pureza genética como um dos principais requisitos para a comercialização da semente, uma vez que a presença de indivíduos atípicos pode resultar em redução da produção, menor uniformidade e menor qualidade do produto final (CASTRO, 2006; BONOW; PINHO; SOARES, 2007; CHOER; BARBIERI; FERREIRA et al., 2011).

A determinação da pureza genética é complexa, pois frequentemente as contaminações não provocam elevadas alterações fenotípicas, especialmente em plantas de polinização cruzada, para isso, é necessária ciência, tecnologia, conhecimento, experiência, bom gerenciamento e comprometimento com a qualidade (CHEDIAK; BRONDANI; DEL PELOSO; MELO; BRONDANI, 2007).

Segundo Ramos (2004), a contaminação varietal pode ser eficientemente controlada nas gerações subsequentes por meio de “rouging” e limpeza de máquinas e equipamentos. Entretanto, podem ocorrer falhas nesse controle que deverão ser detectadas por meio de testes para avaliar a presença de mistura na amostra e a porcentagem de sua incidência.

Dessa forma, a análise da pureza genética, como rotina em programas de melhoramento, na manutenção de sementes do melhorista e em laboratórios de sementes, é indispensável, pois podem ocorrer falhas no programa de controle de qualidade. Para tanto, podem ser usados descritores morfológicos, fisiológicos, citológicos e bioquímicos (AOSA, 1991, BRASIL, 2011 e ISTA, 1996).

No programa brasileiro de controle de qualidade, a pureza genética é identificada segundo as regras para análise de pureza da “Association of Official Seed Analysts” (AOSA, 1983), que também estabelece que diante da impossibilidade de identificação da cultivar por meio de características das sementes, devem ser consideradas características das plântulas ou das plantas em crescimento mediante a observação no mínimo, um mês após a semeadura, para que, através do seu crescimento e desenvolvimento, ocorra sua identificação morfológica.

Vale salientar que essas análises podem ser ineficientes para distinguir genótipos morfológicamente semelhantes (RAMOS, 2004).

Técnicas de marcadores moleculares também são empregadas para esta classificação e identificação (FERREIRA, 2003; VIDAL et al., 2005). Estas são altamente confiáveis, no entanto, possuem limitações para implantação em escala de execução devido principalmente a necessidade de considerável período para execução, serem laboriosas, necessitar de agentes químicos, inviabilizar a semente para outros testes e necessitar de pessoal treinado.

Com isso a busca de novas metodologias analíticas que permitam sua repetitividade a um baixo custo de execução, não invasivas, não destrutivas, não poluentes e rápidas, vem ganhando expressão no mundo o que pode tornar-se uma opção para classificação de sementes de mamona conforme seu genótipo, mantendo as sementes viáveis para outras avaliações.

Desta forma, neste trabalho objetivou-se o desenvolvimento de métodos analíticos de identificação baseados no uso da espectrometria de reflectância no infravermelho próximo (NIR) e das técnicas quimiométricas de cinco genótipos de mamona e cinco gerações de um cruzamento recíproco.

2. FUNDAMENTAÇÃO TEORICA

2.1. A produção de sementes de mamona

A produção de sementes de alta qualidade de *Ricinus communis* L., constitui um grande desafio para a pesquisa agrícola e para toda a cadeia produtiva da cultura, necessitando que sejam adotadas tecnologias que permitam avaliar a qualidade física, genética e fisiológica das sementes.

De acordo com Freire *et al.*, (2001) e Azevedo e Lima (2001), a baixa produtividade média da mamona observada no Brasil deve-se ao uso de sementes de baixa qualidade, multiplicadas pelos próprios agricultores, acarretando alto grau de heterogeneidade e grande diversidade de tipos locais.

Porém, o uso de sementes com potencial fisiológico elevado é fundamental na obtenção de resultados satisfatórios em culturas de expressão econômica (BINOTTI *et al.*, 2008).

2.2. Normas para a produção de sementes de mamona

O Ministério da Agricultura, Pecuária e Abastecimento (Mapa), regulamenta por meio da Instrução Normativa nº45, de 17 de setembro de 2013, os padrões de identidade e qualidade para a produção e a comercialização de sementes de mamona, dentre as quais estabelece parâmetros de campo (**Tabela 1**) e da semente (**Tabela 2**).

A produção de sementes se faz com a escolha de áreas para a instalação de campos de produção que devem ser vistoriadas na escolha e depois da semeadura, com o objetivo de evitar a presença de contaminantes oriundos de outras espécies ou outras cultivares e para eliminar os riscos de se obter uma semente com baixa qualidade e identidade.

Para atingir esse objetivo as vistorias são necessárias, cuja execução exige a aplicação de métodos, processos, técnicas e procedimentos padronizados. No caso da mamona, a implantação criteriosa de um campo de produção de sementes, visa garantir a inexistência de plantas indesejáveis, eliminando o risco de infestação de pragas de fácil disseminação e de elevada importância econômica para a cultura.

Pelo processo de vistorias de campos busca-se controlar as misturas de cultivares que, se presentes em níveis de risco, ocasionariam prejuízos aos usuários de sementes e agricultores

em decorrência da perda de produtividade ou do elevado custo de produção em face da necessidade de uso de agroquímicos em elevadas quantidades para o manejo de pragas.

Tabela 1. Parâmetros de Campo utilizados na produção de sementes de mamona.

Parâmetros de Campo					
		Básica	C1	C2	S1e S2
Vistorias					
Área máxima da gleba (ha)		50	50	50	100
Número mínimo de vistorias		2	2	2	2
Número mínimo de subamostras		200	150	125	75
Rotação (ciclo agrícola)		-	-	-	-
População da amostra		1200	900	750	450
Isolamento (metros)		1000	1000	1000	1000
Plantas atípicas (fora de tipo) (nº máximo)		0/1200	3/900	3/750	3/450
Plantas de outras espécies					
-	Cultivadas/Silvestres/Nocivas Toleradas	-	-	-	-
-	Nocivas Proibidas	-	-	-	-
Pragas (nº máximo de plantas)					
-	Fusariose (<i>Fusarium oxysporum</i> f. sp. <i>ricini</i>)	0/1200	0/900	0/750	0/450
-	Murcha bacteriana (<i>Pseudomonas solanacearum</i>)	0/1200	0/900	0/750	0/450
-	Mofo cinzento do cacho (<i>Botrytes ricini</i>)	6/1200	6/900	7/750	6/450

Vale salientar que em se tratando da mamona existem cultivares distintas cujas sementes são bastante semelhantes em termos de cor, de tamanho e formato e muitas vezes não são percebidas visualmente. Mesmo após a emergência das plântulas, as mesmas, apresentam características semelhante que só podem ser diferenciadas muitas vezes após cerca de 30 dias da emergência, o que promove considerável dificuldade na determinação da pureza varietal de um lote de sementes.

Tabela 2. Parâmetros de semente utilizados na produção de sementes de mamona.

Parâmetros de Semente					
		Básica	C1	C2	S1e S2
Pureza					
-	Semente pura (% mínima)	98,0	98,0	98,0	98,0
-	Material inerte (%)	-	-	-	-
-	Outras sementes (% máxima)	0,0	0,1	0,1	0,1
Determinação de outras sementes por número (nº máximo)					
-	Semente de outra espécie cultivada	0	1	2	2
-	Semente Silvestre	0	1	2	2
-	Semente nociva tolerada	0	1	2	2
-	Semente nociva proibida	0	0	0	0
Germinação (%) mínima		70	80	80	80
Validade do teste de germinação (máxima em meses)		8	8	8	8
Validade do teste de reanálise do teste de germinação		5	5	5	5

A pureza varietal ou genética mínima das sementes deve ser garantida pelo produtor de sementes e qualquer mistura entre materiais distintos, seja de forma intencional ou acidental, pode comprometer diretamente o rendimento esperado em uma plantação e será punido mediante norma do Ministério da Agricultura (BRASIL, 2011).

2.3. Classificação de sementes de mamona

As sementes de mamona possuem características fenotípicas diferentes, como: cores, tamanhos, formatos e massa. A classificação dessas sementes de acordo com o genótipo ou cultivar torna-se necessário, pois denota características específicas como: teor óleo, resistência a pragas, teor de ricina, entre outros, (BELTRÃO; AZEVEDO, 2007; SAVI FILHO, 2005)

Apesar de ter uma variabilidade intrínseca, não é possível fazer uma identificação de alguns genótipos ou cultivares, pois existem sementes semelhantes de cultivares ou linhagens, o que impossibilita por meio da inspeção visual, a identificação correta de algumas delas. Em geral,

o procedimento utilizado no processo de identificação da cultivar é realizado por meio de descritores morfológicos da planta. Para tanto, é realizado o plantio com amostras de sementes sendo aguardado no mínimo um mês, para que, por meio do seu crescimento e desenvolvimento as características fenotípicas da cultivar sejam expressas (MILANI *et al.*, 2009).

Em metodologias alternativas, pode-se ainda empregar técnicas de biologia molecular com marcadores moleculares obtidos da estrutura do DNA.

Contudo, tais metodologias realizam ensaios de identificação lentos, laboriosos, usa reagentes e equipamentos caros, e são difíceis de serem implantadas em escala de rotina, inviabilizando a semente para uso futuro.

A busca de novas técnicas que possam ser implantados em escala de repetições sem promover a destruição, ou mesmo, sem a necessidade que as sementes sejam semeadas para posterior identificação das cultivares ou linhagens baseadas nas características morfológicas da planta, correspondem as técnicas mais desejadas na atualidade e dessa forma surge a espectrometria como técnica promissora.

2.4. Espectroscopia na região do visível

A luz possui muito mais informações do que os nossos olhos são capazes de detectar, essas informações, são baseadas em pequenas e quase imperceptíveis alterações de cores que com o uso de detectores específicos podem gerar informações importantes do material alvo da análise.

A região do visível corresponde a faixa de absorção molecular do espectro que varia de 400 a aproximadamente 750nm, e cuja característica é baseada nos cromóforos permitem transições eletrônicas devido a absorção de energia quantizada, gerando informações ligadas a cor. Mediante as frações de absorbância, transmitância, reflectância e transflectância do visível correspondentes a uma faixa específica são associadas as cores de luz absorvida e luz transmitida (**Tabela 3**) SKOOG *et al.*, 2014

Tabela 3. Frações de absorbância e transmitância específicas e suas correspondentes nas faixas espectrais pertencentes a região do visível.

Faixa de absorbância (nm)	Cor da luz absorvida	Cor complementar transmitida
400-435	Violeta	Amarelo-Verde
435-480	Azul	Amarelo
480-490	Azul-verde	Laranja
490-500	Verde-Azul	Vermelho
500-560	Verde	Roxo
560-580	Amarelo-Verde	Violeta
580-595	Amarelo	Azul
595-650	Laranja	Azul-verde
650-750	Vermelho	Verde-azul

FONTE: SKOOG et al., 2014

Na química analítica a técnica vem sendo cada vez mais utilizada nas avaliações por via úmida, com recente destaque para aplicações em amostras sólidas devido ao avanço dos equipamentos de reflectância (ABDULLATEEF; OSMAN, 2011). Destaca-se ainda quando comparada a outras técnicas espectrométricas, por necessitar de uma instrumentação de custo relativamente baixo frente as outras técnicas (NETO JUNIOR, 2008).

2.5. Espectrometria no infravermelho

A espectroscopia consiste num método analítico, no qual se pode estudar as interações de radiações electromagnéticas com as moléculas (ALCÁCER, 2007). A interação entre dois átomos nas moléculas corresponde a diferentes tipos de energia, como a electrónica, rotacional, translacional e vibracional (COATES, 2000).

De acordo com Reich (2005), a região do infravermelho encontra-se entre as gamas do visível e as microondas no espectro electromagnético, dividindo-se em próximos, médios e longínquos ou distante ($10 - 400\text{cm}^{-1}$) (**Tabela 4**).

De acordo com a The American Society of Testing and Materials (ASTM), verifica-se, que a absorção da radiação infravermelha próxima compreende o intervalo entre 780 – 2526nm (SIESLER et al., 2002; REICH 2005; SKOOG; HOLLER; NIEMAN, 2009).

Tabela 4. Limites aproximados das três regiões do infravermelho.

Região	Comprimento de onda (λ), μm	Número de onda (ν), cm^{-1}	Frequência (ν), Hz
Próximo	0,78-2,5	12800-4000	$3,8 \times 10^{14} - 1,2 \times 10^{14}$
Médio	2,5 - 50	4000 - 200	$1,2 \times 10^{14} - 6,0 \times 10^{14}$
Distante	50 - 1000	200 - 10	$6,0 \times 10^{14} - 3,0 \times 10^{14}$

FONTE: Adaptado de SKOOG; HOLLER; CROUCH, 2009

Para ocorrer absorção por parte de uma molécula, é necessário que a radiação incidente na mesma, provoque excitação vibracional, de modo que a energia corresponda à diferença de dois estados energéticos vibracionais, ou seja, igual à diferença entre estados excitado e fundamental, ou que a frequência de vibração produza alteração do momento dipolar, sejam eles translacionais ou rotacionais. Portanto, é uma técnica que se limita para espécies moleculares com pequenas diferenças de energia entre diversos estados vibracionais e rotacionais (BARBOSA, 2008; HIGSON, 2009; SKOOG; HOLLER; CROUCH, 2009; SKOOG; HOLLER; NIEMAN, 2009; EWING, 2011)

As moléculas necessitam de ser diatômicas heteronucleadas ou poliatômicas para poderem ser excitadas por processo vibracional para que, se verifique, alteração do seu momento dipolar quando ocorre a absorção de uma radiação incidente por parte de uma molécula (SIESLER et al., 2002; ALCACÉR 2007; ALMEIDA, 2009).

Verifica-se, que ao absorver radiação de infravermelhos, haverá uma transição entre os níveis de energia vibracionais, que mediante a frequência e quantidade da mesma, pode indicar qual o tipo e número de ligações entre átomos e grupos funcionais estão presentes (SANTOS, 2009).

A espectroscopia infravermelha é usada pelos químicos orgânicos para ajudar na identificação estrutural, pois sabe-se que cada movimento envolve oscilação em uma frequência específica; logo, a absorção pode ajudar a caracterizar o tipo de oscilação molecular que está ocorrendo e, frequentemente, os grupos funcionais presentes nas moléculas (HIGSON, 2009).

Moléculas diatômicas heteronucleares como, por exemplo, o HCl, o CO₂ e o SO₂, possuem um momento de dipolo significativo, isto é, modos vibracionais de absorção ativos no infravermelho. O contrário ocorre em espécies homonucleares, como O₂, N₂ ou Cl₂, as quais não possuem variação no momento de dipolo tendo, como consequência, a não absorção da

radiação infravermelha (BARBOSA, 2008; SILVA RODRIGUES; EUSÉBIO, 2008; SKOOG; HOLLER; NIEMAN, 2009; EWING, 2011).

Portanto, a absorção da radiação IV está diretamente relacionada à concentração de ligações (ou grupos funcionais) específica na amostra e, portanto, à concentração de analitos existentes (HIGSON, 2009).

Os grupos funcionais cuja absorção mais se salientam são respectivamente $-\text{CH}$, $-\text{NH}$, $-\text{OH}$ e $-\text{SH}$. Estes podem facilitar a determinação de umidade, gorduras, proteínas e hidratos de carbono (REICH 2005; ALMEIDA, 2009; SANTOS 2009).

2.5.1. Vibrações moleculares

Em uma molécula, os átomos se movimentam, flexionam e vibram em torno de suas ligações moleculares. A frequência de uma oscilação molecular é determinada, principalmente, pela natureza de cada ligação na molécula e pelos tipos de grupamento diretamente associados à oscilação (HIGSON, 2009; SKOOG; HOLLER; CROUCH, 2009).

As oscilações ou vibrações moleculares podem ocorrer por: estiramento, que consiste em uma variação interatômica ao longo do eixo da ligação entre dois átomos ou grupos da molécula; bem como por deformação, que envolvem variação de ângulos entre as ligações, além de uma combinação de ambos. Dos modos de vibração conhecidos, os mais utilizados são: a deformação angular simétrica no plano, a deformação angular assimétrica no plano, deformação angular simétrica fora do plano e deformação angular assimétrica fora do plano, como ilustrado na **figura 1**.

Deformações Axiais

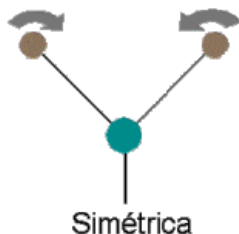


Simétrica

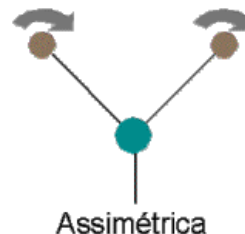


Assimétrica

Deformações Angulares



Simétrica



Assimétrica



Simétrica Fora do Plano



Assimétrica Fora do Plano

Figura 1- Oscilações moleculares de deformação (HIGSON, 2009).

A maioria das moléculas se encontra no estado vibracional fundamental à temperatura ambiente, podendo ocorrer transições entre os níveis vibracionais quando energia é transferida para a molécula. O modelo do oscilador harmônico simples pode ser utilizado para descrever as características de uma vibração atômica, se uma molécula diatômica for aproximada por duas massas esféricas, m_1 e m_2 , conectadas por uma mola que possui constante de força k (PASQUINI, 2003; SKOOG; HOLLER; CROUCH, 2009).

A energia potencial do sistema massa-mola (E), em função da distância entre os átomos (x), tem o formato de uma parábola, como representado pela **figura 2**. A energia potencial é mínima na posição de equilíbrio e máxima quando a mola está esticada ou comprimida na sua amplitude máxima (PASQUINI, 2003). Entretanto, o modelo harmônico possui limitações quando utilizado para descrever os movimentos vibracionais moleculares, pois os sistemas moleculares possuem energia discreta e não energia contínua, conforme descrito pelo modelo. Além disso, nesse modelo a diferença entre dois modos vibracionais adjacentes são sempre iguais, permitindo apenas a transição entre níveis vibracionais adjacentes (PASQUINI, 2003).

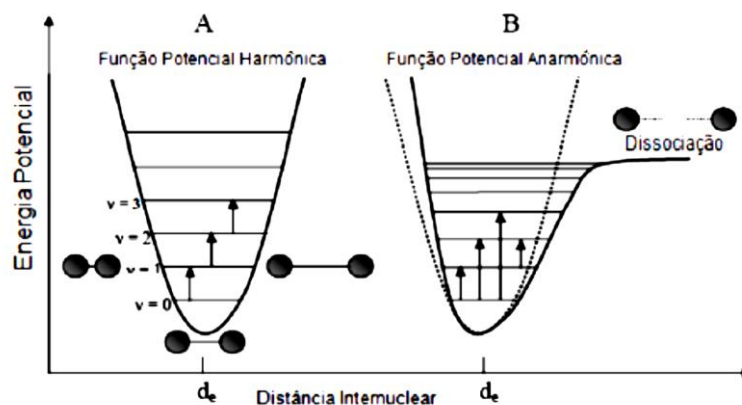


Figura 2- Representação sistemática da energia potencial para (a) oscilação harmônica, (b) oscilador anarmônico. d_e = distância do equilíbrio (E = mínimo). FONTE: Adaptado de PASQUINI, 2003.

Um modelo mais realista para representar as vibrações moleculares é o oscilador anarmônico, que considera a repulsão coulômbica entre dois núcleos, quando os dois átomos se aproximam e o decréscimo da energia potencial, quando a distância interatômica está próximo da dissociação, como mostra a **figura 2-b**. Nesse modelo, é possível explicar transições denominadas sobretons, com frequências aproximadamente duas a três vezes a frequência fundamental (CHAGAS, 2006; SKOOG; HOLLER; CROUCH, 2009).

Devido às vibrações e rotações de diferentes tipos que ocorrem nas ligações da molécula, as posições relativas aos átomos, não são fixas, mas oscilam continuamente; assim, essas vibrações podem ser classificadas em estiramento e deformação (SKOOG; HOLLER; NIEMAN, 2009).

A característica da vibração de estiramento é a variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos; já as vibrações de deformação envolvem uma variação no ângulo entre duas ligações e são de quatro tipos: deformação simétrica no plano, deformação assimétrica no plano, deformação simétrica fora do plano e deformação assimétrica fora do plano (SKOOG; HOLLER; NIEMAN, 2009).

Em temperatura ambiente a maioria das moléculas se encontra no nível vibracional fundamental e as transições permitidas, são denominadas transição fundamental ou 1º harmônico em que este domina o espectro de absorção do infravermelho (SCAFI, 2000; 2005).

Embora o modelo harmônico possa ser útil para entender a espectroscopia vibracional, este modelo não consegue explicar o comportamento de moléculas reais. A principal limitação

é não considerar as forças coulômbicas de atração e repulsão nem a dissociação da ligação (NUNES, 2008)

2.6. Espectrometria no infravermelho próximo – NIR

2.6.1. A espectroscopia NIR

A origem da Espectrometria no Infravermelho data do início do século XIX, com o trabalho pioneiro do músico e astrônomo alemão Frederick Willian Herschel. Em 1800, durante a execução de um experimento que consistia em utilizar um prisma para separar as faixas espectrais associadas a região espectral do visível, Herschel observou que cada uma das cores decompostas pelo prisma apresentava uma quantidade de energia distinta. Para monitorar a quantidade de energia associada a cada cor, ele teve a ideia de utilizar um termômetro, para sua surpresa, ele verificou que acima do vermelho, onde não havia luz visível, era a região que apresentava a maior temperatura (MILAGRES, 2009).

Segundo Nisgoski (2005) os primeiros instrumentos comerciais apareceram apenas em meados do século XX, impulsionados pelo seu desenvolvimento durante a Segunda Guerra Mundial, quando a espectroscopia de infravermelho foi utilizada para o controle da concentração e grau de pureza do butadieno utilizado para fabricação de borrachas sintéticas.

De acordo com Pasquini (2003), as primeiras contribuições do Brasil para o campo de aplicação analítica do NIR foram as publicações de Nishiyama e colaboradores em 1991, nessa mesma década a aparição no mercado de espectrômetros com Transformada de Fourier, resultou em um incremento da expansão desta técnica, vislumbrando-se o surgimento de um dos mais versáteis métodos de análise química de substâncias.

A técnica tem se expandido especialmente por fornecer medidas precisas, possibilitando o monitoramento e o controle de processos. Por isso, atualmente, a espectroscopia é alvo de muitos esforços e grandes investimentos no sentido de buscar o desenvolvimento de aplicações que resultem em informações concisas sobre processos em tempo real (HEIN, 2009).

Baseada na espectroscopia vibracional a técnica NIR mede a interação da luz com o material, que é determinada pela vibração das ligações químicas dos constituintes da amostra, sendo fundamental na absorção seletiva de luz pelos compostos químicos (HEIN, 2009).

A espectroscopia NIR se baseia nas ligações químicas anarmônicas com C e H, principalmente, isso devido as substâncias possuírem frequências de vibração específicas, que correspondem a níveis de energia das moléculas e estas, são dependentes da forma, da superfície de energia potencial da molécula, de sua geometria e da massa dos átomos envolvidos no seu acoplamento vibracional.

A intensidade das bandas de absorção no NIR é cerca de 10 a 1000 vezes mais fracas que sua banda fundamental na região do infravermelho médio (MIR). Isto poderia a princípio, ocasionar uma desvantagem devido à diminuição da sensibilidade analítica (LIMA et al., 2009), entretanto, tal dificuldade pode ser superada com o uso de fontes de radiação interna e detectores de alta eficiência que contribuem para o aumento da relação sinal/ ruído (HONORATO, 2006).

A espectroscopia NIR comporta a região de números de onda adjacentes à MIR e se estende até a região do visível. A absorção NIR é baseada em sobretons (*overtone*) e combinação das vibrações da molécula investigada e suas baixas probabilidades de transição, uma vez que as intensidades de absorção decrescem sucessivamente na direção do MIR ao visível. Existe a possibilidade de ajuste na espessura da amostra, passando da ordem de grandeza de milímetros até centímetros, dependendo da classificação do sobretom (SIESLER et al., 2002).

A energia de radiação na faixa do NIR, quando absorvida, é transformada em energia mecânica (vibracional) associada ao movimento dos átomos unidos por ligações químicas moleculares as quatro regiões do NIR contém “ecos” da absorção fundamental (PASQUINI, 2003) e portam informações complexas de serem extraídas sem o auxílio das ferramentas estatísticas multivariadas (CAMPOS, 2008). A vibração dos átomos no interior de uma molécula apresenta energia coerente com a vibração do espectro eletromagnético correspondente ao infravermelho (100 a 10.000 cm^{-1}) (LUZ, 2003).

Quando uma molécula absorve radiação na região do infravermelho, ocorrem alterações no estado energético das ligações. A energia de um raio luminoso incidente é absorvida quando a frequência da luz é igual a frequência da ligação. Assim, o espectro é obtido por meio do registro da intensidade absorvida em função do comprimento de onda da luz incidente (HEIN, 2009).

Diversas são as vantagens de utilização da metodologia NIR e MIR para caracterização de processos. A **tabela 5** apresenta uma comparação entre as características qualitativas das técnicas de espectroscopia vibracional MIR e NIR. A partir desta tabela é possível escolher, baseado nas características almejadas no processo de caracterização, qual das técnicas apresenta melhor viabilidade para aplicação.

Tabela 5. Limites aproximados das duas regiões do infravermelho MIR e NIR.

	MIR Vibrações fundamentais	NIR Sobretons e Combinações
Qualitativa	Excelente (estrutura)	Ruim (identidade)
Quantitativa	Excelente	Excelente
Intensidade	Alta	Baixa
Espessura da amostra	Muito pequena	Grande
Matérias	KBr/NaCl	Quartzo/Vidro
Sinal/Ruído	$<10^4$	$>10^4$
Reflectância	Satisfatória	Excelente

FONTE: Adaptado de Pasquini (2002).

Dentre as vantagens que motivam o uso da espectroscopia no infravermelho próximo (*NIRS–Near Infrared Spectroscopy* ou simplesmente NIR – *Near Infrared*) na caracterização de processos, podem ser citadas a velocidade de medida quando comparada a outros processos (usualmente medidas são tomadas em menos de 1 segundo), ser uma técnica não destrutiva e que na grande maioria dos casos não necessita de preparação da amostra, além de ser considerada altamente versátil. Caso as amostras possuam ligações C-N, N-H ou O-H, e caso a concentração do analito ultrapasse 1% da composição total, resultados promissores podem ser esperados quase que na totalidade dos casos (PASQUINI, 2003; SKOOG; HOLLER; NIEMAN, 2009; ALVES; POPPI, 2013).

Vale salientar que a técnica não mede diretamente o constituinte químico. O que a técnica de espectros de NIR mede é a interação da radiação com as ligações químicas dos constituintes da amostra.

Por possuir baixa absorvidade, a espectroscopia NIR permite melhor penetração da radiação em amostras sólidas e análises diretas de fortes absorventes como, por exemplo, de líquidos turvos ou sólidos nos modos de reflectância, transmitância ou transflectância, conforme a **figura 3**, sem necessidade de pré-tratamento da amostra (SIMÕES, 2008; LIRA, 2010).

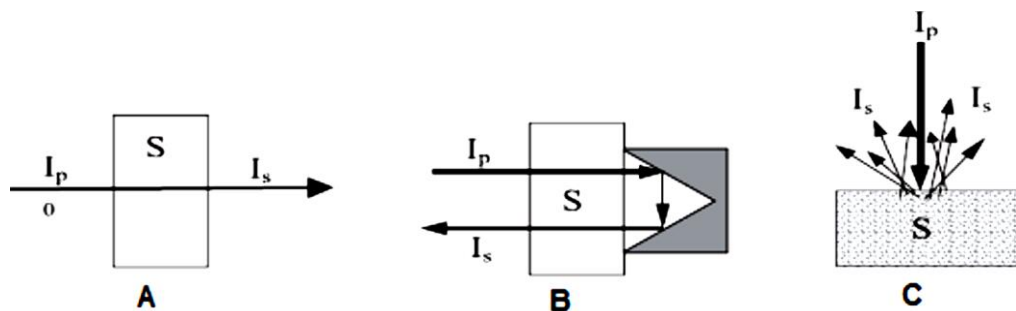


Figura 3- Modos de medição utilizados em espectroscopia NIR. 11 (A) transmitância; 11 (B) transflectância e 11 (C) reflectância difusa, através do meio de dispersão (PASQUINI, 2003).

Na **figura 3 (A)** observa-se o modo de transmitância, muito usado na espectrometria UV – VIS convencional. As amostras são medidas em cubetas de vidro ou quartzo com percurso óptico variando de 1 a 50mm.

O modelo de transflectância é representado na **figura 3 (B)**, para esse tipo de medida usam-se feixes de fibra óptica ou dispositivos para este fim, diferenciando-se das medidas de transmitância pelo caminho óptico duplo.

As medidas de reflectância difusa de amostras sólidas **figura 3 (C)** formam a base das medidas NIR, com predominância dos fenômenos de espalhamento e absorção de partículas sólidas.

As absorções moleculares mais frequentes na espectroscopia NIR são anarmônicas de estiramentos C-H, N-H e O-H e bandas de combinação (PASQUINI, 2003; HIGSON, 2009; SKOOG; HOLLER; CROUCH, 2009). Essa região é caracterizada por possuir bandas sobrepostas e fracas, aproximadamente de 10 a 100 vezes mais fracas quando comparadas às bandas do MID, visto que as probabilidades de transições envolvendo sobretons são menores quando comparadas às transições fundamentais (COSTA FILHO, 2003).

Estruturalmente, os equipamentos de espectroscopia NIR são similares aos instrumentos para medidas de espectroscopia UV-visível e MIR sendo possível encontrar espectrofotômetros UV-VIS comerciais projetados para operar na região de 180 a 2500 nm possibilitando a obtenção de espectros NIR. Os componentes base são a fonte luminosa onde se utiliza lâmpadas de tungstênio-halogênio como fonte, detectores fotocondutores de sulfeto de chumbo (PbS) e seleneto de chumbo (PbSe), a fotodiodos de antimoneto de índio (InSb) e arseneto de índio (InAs) e células ópticas de quartzo ou de sílica fundida, além de detector e o elemento dispersivo, que pode ser um prisma ou então uma rede de difração (HIGSON, 2009; SKOOG; HOLLER; CROUCH, 2009).

Assim como todo método analítico de caracterização, a técnica de espectroscopia NIR também está sujeita à presença de ruídos nas medidas. De forma genérica, toda medida analítica é composta pelo sinal, que contém a informação qualitativa e quantitativa desejada, e o ruído, que nada mais é que uma fonte de informação estranha que afeta o limite de detecção, precisão e exatidão do método. Em NIR, a fonte de ruído pode ser oriunda de duas vertentes, o ruído químico e o ruído instrumental.

O ruído químico resulta de uma grande quantidade de variáveis que afetam os sistemas químicos, principalmente variações físicas e físico-químicas não detectadas como temperatura, tamanho e distribuição de partículas, polimorfismo, pressão, densidade, textura, forma, compactação, pH, umidade, dentre outras (COSTA FILHO, 2003).

O ruído instrumental por sua vez, é associado aos componentes do equipamento, como fontes, transdutores, processadores de sinal, dentre outras. Desta forma, para elevar o nível das medidas em sistemas de espectroscopia NIR, assim como nas demais técnicas analíticas, é aconselhada a utilização de elementos de hardware como filtros, moduladores e detectores sincronizados associados a sistemas de software capazes de permitir a extração de sinais ambientais que impõe ruídos as medidas, principalmente filtros digitais.

2.7. Região infravermelho próximo de ondas curtas-SWNIR

Na região do IR existem duas faixas que são muito utilizadas em espectroscopia: o infravermelho próximo de ondas curtas (*short wave near infrared*, SWNIR) (780-1100nm) e o infravermelho próximo (*near infrared*, NIR) (1100-2500nm) (PAVIA, 2001).

De acordo com Barboza (2003) a região do espectro compreendida entre as faixas de 780 a 1100nm recebeu a nomenclatura de região do infravermelho próximo de ondas curtas (SW-NIR) ou região de Herchel que teve a divisão baseada no tipo de detector utilizado para as medidas.

A região SW-NIR é utilizada para medir vibrações moleculares e bandas de combinação de grupos C-H (gorduras, óleos, hidratos de carbono), segundos sobretons (tons de frequência superior à da frequência fundamental) de grupos O-H (água, álcool) e N-H (proteínas) e terceiros sobretons de grupos C-H. A posição exata das bandas depende do ambiente químico, havendo um grau elevado de exclusividade de espectros para as diferentes moléculas orgânicas. As vantagens do SWNIR incluem um valor elevado da razão sinal/ruído (tipicamente 25000:1) assim como um alto rendimento utilizando cabos de fibra ótica, pois este tipo de cabos orienta um feixe de luz a partir de um ponto para outro sem que ocorram efeitos de dispersão, reflexão

e refração. Portanto, a sua utilização com técnicas espectroscópicas é altamente atrativa para monitorização de bioprocessos (WORKMAN, 2000).

2.8. Avaliações das bandas espectrais

A espectroscopia na região do infravermelho próximo é caracterizada pelos sobretons e bandas de combinação em virtude da anarmonicidade das frequências de vibração das ligações envolvidas na absorção mas, a interpretação e o conhecimento sobre as bandas espectrais nessa região do espectro são fatores importantes para o sucesso da tecnologia NIR.

O diagrama apresentado na **figura 4** contém as correlações entre a absorção dos grupos funcionais e o comprimento de onda observadas na região do NIR (CHAGAS, 2006).

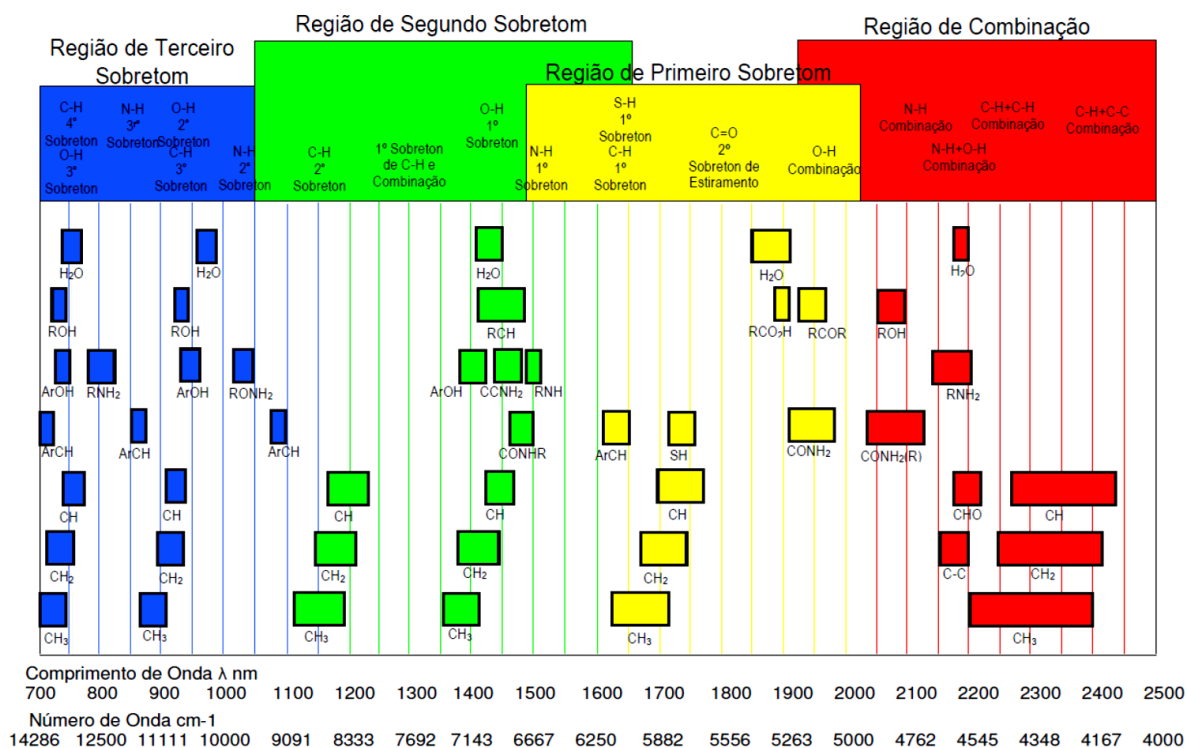


Figura 4- Principais bandas de absorção (Frequência de grupos funcionais) na região NIR e sua localização no espectro eletromagnético. FONTE: Adaptado de XIAOBO et al., 2010

2.9. Principais aplicações do NIR

A utilização da espectrometria no infravermelho vem ocupando espaço na indústria e em diversos setores produtivos, assumindo função de destaque principalmente em análises quantitativas e qualitativas de produtos e processos. Estas análises são possíveis porque a espectrometria no infravermelho tem como resultado espectros que se relacionam com algumas características das substâncias analisadas e do ambiente em que estão inseridas, podendo assim responder a uma série de variáveis de produto e processo.

Foi na agroindústria que a espectrometria NIR conheceu suas primeiras aplicações, tornando-se bastante difundida nesta área. Basicamente, as aplicações da técnica se davam na quantificação de substâncias como água, óleo, gorduras e proteínas em produtos agrícolas (WATSON, 1977).

Grande parte destes estudos foi baseada nos trabalhos da década de 60 do engenheiro agrícola norte-americano Karl Norris, o qual é geralmente citado na literatura como “o pai” da espectrometria NIR moderna. Em seguida, a técnica se estendeu à química orgânica, indústria de polímeros e indústria farmacêutica, além de aplicações crescentes nas áreas de petróleo, biomédica, dentre outras (HOMMA et al., 1996; KIM et al., 2000)

Em particular, a técnica foi utilizada recentemente na detecção de defeitos e fungos em grãos de café, em substituição à técnica visual, na verificação de propriedades moleculares tamanho de partículas, processo de reforma catalítica e caracterizações de produtos farmacêuticos (SANTOS, 2000; SCAFI, 2000; MARTINS; MUNIZ, 2002; YUAN et al., 2003; PARIZZI et al., 2005).

Com o desenvolvimento recente da miniaturização das partes essenciais de instrumentos NIR, tem sido possível a construção de equipamentos portáteis, ideais para análises de campo, o que permitirá ampliar ainda mais as aplicações da espectrometria NIR como sensor para os mais diversos fins (ARAÚJO, 2007). Trabalhos têm sido desenvolvidos recentemente considerando o uso da espectrometria de infravermelho próximo em processos químicos, seja na indústria ou em trabalhos acadêmicos. Na maioria dos casos, o instrumento é acoplado a um microcomputador que implementa aquisições automáticas e aplica modelos matemáticos aos espectros coletados. Em geral, o intervalo entre a tomada de um espectro e o cálculo do componente de interesse se dá em torno de um minuto, ou menos, a depender da configuração do equipamento e do grau de precisão desejado, baseado no número de varreduras que são efetuadas. Os dados podem ser transferidos para computadores centrais, permitindo então a tomada de decisão em tempo real (RODRIGUES, 2012).

A aplicabilidade da espectroscopia NIR é bem diversificada, sendo utilizada em diversos segmentos, tais como na área agrícola (SENA; POPPI, 2000; MORGANO et al., 2005;), alimentícia (BOTELHO; MENDES; SENA, 2013), indústrias petroquímicas (PANTOJA, 2010) farmacêutica (SIMÕES, 2008) e no monitoramento de reações orgânicas (KILLNER; ROHWEDDER; PASQUINI, 2011).

A grande aceitação da espectroscopia NIR por diferentes tipos de indústria pode ser atribuída a grande facilidade de se trabalhar com amostras sólidas e em forma de pó, realizando pouca ou nenhuma manipulação analítica, permitir a análise simultânea de vários parâmetros, as análises são não destrutivas e são pouco invasivas, rápidas, de baixo custo, confiáveis e versáteis e não consomem reagentes químicos (NAES et al., 2002; SIMÕES, 2008).

2.10. Quimiometria

A quimiometria trata da aplicação de ferramentas matemáticas e estatísticas para o planejamento, otimização e extração de informações partindo de um conjunto de dados físico-químicos multivariados buscando encontrar as relações entre variáveis mensuradas e amostras, convertendo essas informações em vetores latentes (KUMAR et al., 2014).

Introduzida no final dos anos 60, por grupos de pesquisa das áreas de química analítica e físico-química orgânica, seu desenvolvimento foi devido à disponibilidade de equipamentos de análise com respostas multivariadas, assim como a disponibilidade de microprocessadores com elevada capacidade computacional, o que permitiu o desenvolvimento de medidas capazes de tratar grande quantidade de informações de forma simultânea (GELADI, 2003).

A técnica se propõe a solucionar problemas de interesse e origem na química, ainda que as ferramentas de trabalho provenham principalmente da matemática, estatística e computação envolvendo assim o planejamento e a otimização de experimentos; pré-processamento de dados espectrais; reconhecimento de padrões; seleção de variáveis e amostras; calibração multivariada e transferência de calibração (BEEBE; PELL; SEASHOLTZ, 1998; FERREIRA et al., 1999).

Os métodos de análise exploratória permitem a interpretação multivariada de conjuntos de dados complexos por meio de gráficos bi- ou tridimensionais. Os métodos supervisionados possuem a vantagem de necessitar de informações prévias a respeito das amostras, para possibilitar a formação de classes e desenvolvimento de modelos baseados nas informações contidas nas mesmas. Entretanto, nos métodos não supervisionados, a separação de classes ocorre baseada em similaridades estatísticas dos dados espectrais, sem a necessidade de

informações sobre a natureza das amostras. Dentre os objetivos dos referidos métodos estão a identificação de grupamentos naturais entre as amostras e o reconhecimento de tendências e padrões (BRERETON, 2007).

À análise exploratória e a calibração multivariada tem por objetivo estabelecer modelos capazes de relacionar uma elevada quantidade de medidas (variáveis independentes) químicas, físico-químicas ou espectrais de uma dada amostra com a finalidade de encontrar quais as variáveis que mais afetam determinado processo, bem como as interações entre elas, de forma a determinar as melhores condições de análise (KOWALSKI, 1983). De maneira geral, esses métodos podem ser classificados como métodos supervisionados, nos quais estão enquadrados a Análise discriminante linear (LDA – *Linear Discriminant Analysis*), o Método dos K-vizinhos mais próximos (KNN – *K-Nearest Neighbor*), a Análise Discriminante com Calibração Multivariada por Mínimos Quadrados Parciais (PLS-DA – *Partial Least Square Discriminant Analysis*), ou não supervisionados, como por exemplo, a Análise de Componentes Principais (PCA – *Principal Component Analysis*) e a Análise de Agrupamento Hierárquico (HCA – *Hierarchical Cluster Analysis*) (WEHRENS 2011).

2.10.1. Pré processamento

Na obtenção dos espectros no infravermelho muitas informações são armazenadas de acordo com a amostra utilizada o que propicia grandes possibilidades de serem empregados os mais diversos tipos de análises químicas e/ou físicas. Os dados originais provenientes de técnicas instrumentais podem apresentar alterações não desejadas, como ruídos instrumentais, intensidade com magnitudes diferentes e variação sistemática da linha de base. Essas alterações espectrais, não possuem, normalmente, relação com a composição da amostra e, portanto, não contribuem para os modelos multivariados, o que pode ocasionar distorções ao modelo. Com o intuito de minimizar este problema, utilizam-se diversos métodos de pré-processamentos buscando remover estas distorções (MASSART et al., 1997; FERREIRA; ANTUNES; MELGO; VOLPE, 1999; BUENO, 2011).

A maior contribuição desta variação pode ser atribuída à falta de estabilidade do instrumento, ao espalhamento da radiação durante a realização das medidas ou à variabilidade das propriedades físicas da amostra (BEEBE; PELL; SEASHOLTZ, 1998).

Nas variáveis podem ser aplicadas três técnicas básicas, iniciais de pré-processamento: centralização dos dados na média, o escalonamento e o auto-escalonamento. A centralização

dos dados na média pode ser definida como a subtração dos elementos de cada linha pela média da sua respectiva coluna. No escalonamento cada elemento de uma linha é dividido pelo desvio padrão da sua respectiva variável, fazendo com que todos os eixos da coordenada sejam conduzidos à mesma escala. O auto-escalonamento consiste em centralizar os dados na média e os desvios padrão de cada variável dividido por sua média. Alguns pré-processamentos aplicados às amostras se destacam, tais como: correção de espalhamento multiplicativo (MSC), variação normal padrão (SNV), suavização, derivação, entre outros (MASSART et al., 1997; BEEBE; PELL; SEASHOLTZ, 1998; FERREIRA et al., 1999; BUENO, 2011).

As técnicas mais usadas no pré-processamento de dados aplicadas no domínio das amostras são: normalização, ponderação, suavização e correção da linha de base (BEEBE; PELL; SEASHOLTZ, 1998).

A normalização é efetuada dividindo-se cada variável por uma constante, a partir de uma análise preliminar dos dados. Na ponderação se atribuem as amostras mais importantes, pesos proporcionais, multiplicando-se cada elemento do vetor amostra pelo seu peso. A técnica de suavização de ruído é usada para aumentar a relação sinal/ruído. Com esta finalidade podem ser utilizados os seguintes filtros digitais: Savitzky-Golay (SAVITZKY; GOLAY, 1964; BEEBE; PELL; SEASHOLTZ, 1998), transformada de Fourier (CERQUEIRA; POPPI; KUBOTA, 2000) e transformada Wavelet (GALVÃO et al., 2001).

Destaca-se o método de Savitzky-Golay que ajusta um polinômio de baixa ordem aos pontos de uma janela pelos mínimos quadrados.

As variações sistemáticas não relacionadas com a propriedade de interesse analítico são descritas como feições da linha de base. Elas podem dominar a análise, se não removidas. Para sua correção pode-se usar: derivação e correção multiplicativa de sinais (MSC) e a variação normal (SNV) (BEEBE; PELL; SEASHOLTZ, 1998).

MSC é um método de transformação que busca compensar os efeitos aditivos e/ou multiplicativos em dados de natureza espectral. Tal método minimiza a influência de efeitos físicos nos espectros, tais como o tamanho de partícula, a rugosidade e opacidade, os quais não trazem informações químicas sobre as amostras e introduz variações espectrais, como o deslocamento da linha de base (SOUZA; POPPI, 2012).

Em SNV, um auto-escalonamento no domínio das amostras é realizado, corrigindo os efeitos de espalhamento da radiação e tamanho das partículas. Problemas associados com a variação da linha de base podem ser corrigidos através da 1ª derivada que possibilita, também, uma melhor visualização dos picos existentes nos sinais originais. A filtragem de ruído aleatório

e aumento da relação sinal/ruído pode ser alcançada através da aplicação de procedimentos de suavização dos dados (VILAR, 2014).

2.10.2. Reconhecimento de padrões não supervisionados

As técnicas de reconhecimento de padrões não supervisionados são aquelas usadas para avaliar a existência de similaridade ou diferenças entre os agrupamentos naturais das amostras, não sendo necessário se utilizar o conhecimento prévio dos membros das classes das amostras. Os principais métodos deste tipo são: análise de agrupamento hierárquico (*Hierarchical Cluster Analysis* – HCA) e análise de componentes principais (*Principal Component Analysis* – PCA) (BEEBE; PELL; SEASHOLTZ, 1998; BARROS NETO; SCARMINIO; BRUNS, 2006).

2.10.2.1. Análise de componentes principais (PCA)

A PCA é um dos métodos multivariados mais comuns empregado na análise de dados complexos (BROWN, 1995; FERREIRA, 2002). Ela permite a interpretação multivariada de conjuntos de dados complexos e com grande número de variáveis como, por exemplo, espectros no infravermelho próximo, por meio de gráficos bi ou tridimensionais. Esses gráficos apresentam informações que expressam a existência de correlação linear entre diversas variáveis facilitando a interpretação multivariada do comportamento da amostra (BEEBE; PELL; SEASHOLTZ, 1998; SABIN; FERRÃO; FURTADO, 2004).

A utilização da PCA visa maximizar as informações reduzindo a dimensionalidade do conjunto de dados original ou pré-processados, minimizando a covariância entre as variáveis, sem perda de informações, permitindo a observação de semelhança e diferença entre as amostras. Na redução da dimensão dos dados, a matriz original é representada por novas variáveis que são ortogonais e linearmente independentes entre si. Essas variáveis, chamadas de componentes principais, PC's, são direcionadas em função da distribuição dos dados. Em um gráfico de escores, a primeira PC é o eixo cuja direção descreve a maior variabilidade dos dados, por isso possui a maior variância explicada. A segunda PC descreve a segunda maior variabilidade do sistema e assim por diante (GELADI; KOWALSKI, 1986; MARTENS; NAES, 1989; FERREIRA, 2002).

De acordo Beebe; Pell; Seasholtz (1998); Sena; Frigueto; Valarini, (2000); Sabin; Ferrão; Furtado (2004); Souza; Poppi (2012), na PCA a matriz X é decomposta em um produto de duas matrizes, divididas em *escores* e pesos.

Os escores representam as coordenadas das amostras no sistema de eixos formados pelas componentes principais as quais são ortogonais. Cada PC é constituída pela combinação linear das variáveis originais e os coeficientes da combinação são denominados pesos. Os pesos são os cossenos dos ângulos entre as variáveis originais e as componentes principais (PCs). A PC1 é definida na direção da máxima variação no conjunto de dados, a PC2 é traçada perpendicular à primeira, com o intuito de descrever a maior porcentagem da variação não explicada pela PC1 (resíduo) e assim por diante.

Enquanto os escores representam as relações de similaridade entre as amostras, os pesos representam a contribuição de cada variável para a formação das PCs. Por meio da análise conjunta do gráfico de escores e pesos é possível verificar quais variáveis são responsáveis pelas diferenças observadas entre as amostras. Uma das ferramentas utilizadas para determinar o número de PCs, a ser utilizado no modelo PCA é a porcentagem de variância explicada acumulada (WOLD; ESBENSEN; GELADI, 1987; KAMAL-ELDIN; ANDERSSON, 1997; SABIN; FERRÃO; FURTADO, 2004).

Na **figura 5** observa-se um plano bidimensional para duas variáveis, com duas primeiras componentes principais (PCs), apresentando as coordenadas das amostras com relação as PC's que representam os escores e são denominadas as projeções das amostras no eixo das PC's. As variáveis originais com maior peso na combinação linear da PC1 e PC2 são as de mais importância (MASSART et al., 1997; BEEBE; PELL; SEASHOLTZ, 1998, GREMPERLINE, 2006)

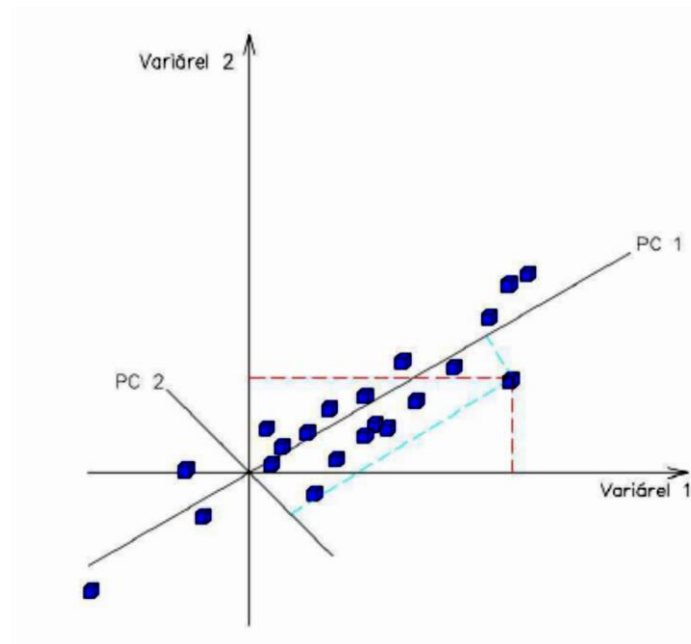


Figura 5- Plano bidimensional com duas variáveis e representação das duas primeiras componentes principais. Adaptado de Beebe et al., 1998.

Utiliza-se a PCA para: identificar amostras anômalas (outliers) com auxílio de testes estatísticos envolvendo loadings e escores, conhecer a contribuição das variáveis, relacionar a contribuição das variáveis originais no novo conjunto de variáveis obtido, entre outros (MOITA NETO; MOITA, 1998)

2.10.3. Reconhecimento de padrões supervisionados

Corresponde a técnica usada para prever se uma amostra desconhecida pertence a uma classe conhecida, a várias classes ou mesmo a nenhuma. Faz-se necessário que haja informações adicionais referentes aos membros das classes, ou seja, é necessário um conjunto de treinamento com objetos de categorias conhecidas para a elaboração de modelos que sejam capazes de identificar amostras desconhecidas. Dentre as técnicas de reconhecimento de padrões supervisionadas podem ser citadas: a modelagem independente e flexível por analogia de classes (*Soft Independent Modeling of Class Analogy* - SIMCA), a análise discriminante dos mínimos quadrados parciais-PLS-DA e a análise discriminante linear (*Linear Discriminant Analysis* - LDA) (BRUNS; FAIGLE, 1985; DERDE; MASSART, 1988; BALLABIO; CONSONI, 2013; OTTO, 1999)

2.10.3.1. Modelagem independente e flexível por analogia de classes (SIMCA)

O método de reconhecimento de padrão supervisionado SIMCA (*Soft Independent Modeling of Class Analogy*), introduzido por Svante Wold em 1976, tem sido um dos métodos mais utilizados na classificação de amostras. Esta ferramenta considera para a classificação, as informações da distribuição da população ou de um conjunto de amostras, estima um grau de confiança da classificação e pode prever novas amostras como pertencentes a uma ou mais classes ou nenhuma classe, tomando para isso o uso da PCA para modelar a forma e a posição do objeto definido pelas amostras no espaço linha visando à definição de uma classe (WOLD, 1976; BRUNS; FAIGLE, 1985; DASZYKOWSKI et al., 2007; CARNEIRO 2008; STUMPE et al., 2012).

Construído a partir de modelos baseados em componentes principais (PC), onde um conjunto de treinamento é delimitado por um número ótimo de componentes principais que é determinado independentemente para cada classe analisada. O modelo final é obtido através da definição da fronteira entre cada PCA modelada. Como consequência, uma hipercaixa é obtida com tantas dimensões quantas forem o número de PC's mais adequados para cada classe. Uma vez definidas as hipercaixas é importante verificar se existem classes com sobreposição (FLATEN; GRUNG; KVALHEIM, 2004; BARROS NETO; SCARMINIO; BRUNS, 2006; FLUMIGNAN, 2010).

O poder de discriminação dos modelos SIMCA se baseia na não existência de sobreposição entre duas classes diferentes, bem como na maior distância interclasses possível. A classificação de uma amostra externa àquelas empregadas na modelagem SIMCA é baseada nas diferentes projeções possíveis no espaço dos escores. Após seu estabelecimento, os modelos são utilizados para classificar amostras futuras como pertencendo a uma das classes. Isto ocorrerá quando a amostra apresentar características semelhantes que a permitam ser inserida neste espaço multidimensional de uma das classes (BEEBE; PELL; SEASHOLTZ, 1998; SCAFI, 2000; SABIN; FERRAO; FURTADO, 2004).

Segundo Sanches (2009), ao se comparar o método PCA com o SIMCA, pode-se dizer que a PCA simplesmente define um sistema de coordenadas no qual as amostras estão representadas, sendo esta utilizada apenas para análise exploratória dos dados, e não para classificação. Já o SIMCA utiliza os eixos definidos por estes sistemas de coordenadas, estabelecendo fronteiras sendo usado justamente para classificar amostras.

O SIMCA baseia-se no cálculo da distância da amostra ao modelo, utilizando-se, para isto, a variância residual para cada amostra da classe X (S_i) (**Equação 1**) e a variância residual total, S_0 (**Equação 2**) (SCAFI, 2000; FLATEN; GRUNG; KVALHEIM, 2004; BRANDEN; HUBERT, 2005; POVIA, 2007).

$$S_i^b = \sqrt{\frac{\sum_{j=1}^M (res_j^b)^2}{M - A_b}} \quad (1)$$

$$S_0^b = \sqrt{\frac{\sum_{i=1}^{N_b} \sum_{j=1}^M (res_{ij}^b)^2}{(N_b - A_b - 1) \cdot (M - A_b)}} \quad (2)$$

Em que:

N_b - número de amostras pertencentes ao conjunto de treinamento da classe b ;

A_b - número de componentes principais utilizados pela classe b ;

M - número de variáveis,

i e j - índices das amostras e variáveis, respectivamente.

A localização da amostra em relação ao modelo é verificada por meio de um teste F , o qual compara o valor obtido pela **equação 3** (F_{cal}) com um valor crítico (F_{crit}) que pode ser obtido empiricamente ou tabelado para determinado nível de confiança. No caso da amostra investigada apresentar um valor de F_{cal} menor que o obtido pelo F_{crit} , esta amostra pertencerá, então, à classe em consideração (WOLD; SJOSTROM, 1977; BLANCO et al., 1998; SCAFI, 2000; FLATEN; GRUNG; KVALHEIM, 2004; POVIA, 2007).

$$F_{cal} = \frac{(S_i^b)^2}{(S_0^b)^2} \cdot \frac{N_b}{N_b - A_b - 1} \quad (3)$$

O SIMCA é um método bem estabelecido na literatura e é largamente utilizado para classificação de amostras em conjuntos de dados com alta dimensionalidade. Alguns estudos descrevem o uso deste método, tais como: Ferreira (2002), Sabin; Ferrão; Furtado (2004), Morales et al., (2008), Sanches (2009), Veras et al., (2011).

De acordo com Beebe; Pell; Seasholtz (1998) a classificação SIMCA pode ser expressa por dois tipos de erro:

- ✓ **Tipo I:** a amostra não é classificada em sua classe verdadeira;
- ✓ **Tipo II:** a amostra é classificada em uma classe distinta da sua.

Com base nesses tipos de erro, uma mesma amostra poderá não ser classificada na sua classe verdadeira e ser ou não classificada em outra(s) classe(s).

2.10.3.2. *Análise discriminante linear (LDA)*

A técnica que utiliza informações das categorias associadas a cada padrão para extrair linearmente as características mais discriminantes. Utilizando a análise discriminante linear podemos também realizar a discriminação entre classes, através de processos supervisionados (quando se conhece o padrão), a LDA é considerada e utilizada como um método supervisionado na concepção, estatística e deve ser empregada quando algumas condições forem atendidas, como: (1) as classes sob investigação são mutuamente exclusivas, (2) cada classe é obtida de uma população normal multivariada, (3) duas medidas não podem ser perfeitamente correlacionadas, entre outras. Consiste em estimar uma combinação linear de duas ou mais variáveis independentes. Ela obtém funções discriminantes lineares as quais maximizam a variância entre as classes e minimizam a variância dentro de cada classe. No caso da existência desta função pode-se dizer que os pontos pertencentes às duas classes são linearmente separáveis (BRUNS; FAIGLE, 1985; MASSART et al., 1997; BALABIN; SAFIEVA, 2008; CASALE et al., 2010; DINIZ et al., 2012).

A LDA se assemelha à PCA, pois ambas buscam reduzir a dimensionalidade dos dados na matriz de variáveis, enquanto a PCA busca encontrar uma direção que tenha a máxima variância dos dados e um mínimo de dimensões relacionadas; a LDA tem a finalidade de selecionar uma direção por meio da qual se alcance a separação máxima entre as classes avaliadas (YU; YANG, 2001; PONTES, 2009).

Mesmo sendo um método de ampla utilização, a LDA possui, algumas desvantagens quando comparada com os outros métodos de reconhecimento de padrões supervisionados, sendo a mesma apropriada apenas para conjuntos de dados de pequenas dimensões. Bem como a sua capacidade de generalização pode ser comprometida por problemas de colinearidade, sendo necessários procedimentos de redução de dimensionalidade e/ou seleção de variáveis. Nesse contexto, aplicações envolvendo técnicas de redução de dimensionalidade têm crescido consideravelmente nos últimos anos (YU; YANG, 2001; GAMBARRA NETO, 2008; SOARES et al., 2013).

2.10.3.3. Algoritmo das projeções sucessivas (SPA)

O objetivo do SPA consiste em buscar um subconjunto representativo pequeno de variáveis espectrais com ênfase na minimização da colinearidade (BREITKREITZ et al., 2003). Sua formulação busca solucionar problemas de classificação e envolve basicamente duas fases. Na primeira, uma sequência de operações de projeções envolvendo as colunas da matriz da resposta instrumental é empregada para formar subconjuntos de variáveis com baixa colinearidade, enquanto que na outra fase, o melhor subconjunto de variáveis é selecionado com base numa função de custo associada ao risco médio de uma classificação incorreta pela LDA (PONTES et al., 2005).

Diversas aplicações do SPA a dados espectroscópicos demonstram seu potencial. Determinação de enxofre em diesel (BREITKREITZ et al., 2003), determinação de parâmetros de qualidade em óleos vegetais (PEREIRA et al., 2008), quantificação de biodiesel em diesel (FERNANDES et al., 2011) e determinação de parâmetros de qualidade de óleos isolantes (PONTES et al., 2011) são alguns exemplos de aplicações do SPA empregando espectroscopia NIR.

Recentemente, Soares et al., (2014) propôs um critério para o SPA-LDA baseado na validação interna. Mais especificamente, o cálculo do custo é realizado não mais com o conjunto de validação, mas sim com o conjunto de treinamento, levando em conta o número de graus de liberdade.

2.10.3.4. Análise discriminante pelos mínimos quadrados parciais (PLS-DA)

Dentre as técnicas de reconhecimento de padrões supervisionados, destaca-se a PLS-DA. Essa é muitas vezes comparada com a análise por componentes principais (PCA) em termos de sua capacidade para classificar dados ou para discriminação entre diferentes grupos. A PCA identifica a variabilidade bruta e não diferencia variações dentro dos grupos e entre os grupos.

Quando a variação entre o grupo domina em relação à variação dentro do grupo, a maior variação encontrada pela PCA será consistente com uma separação entre o grupo. No entanto, se a variação dentro do grupo é maior do que a variação entre o grupo, então PCA não é mais eficaz. Nesses casos, a PLS-DA é uma alternativa preferida à PCA (BARKER; RAYENS, 2003).

A PLS-DA é uma técnica que visa encontrar variáveis latentes no espaço multivariado que discriminem classes no conjunto de treinamento usando um modelo de mínimos quadrados parciais (PLS) e calculando-se um limiar, onde o número de falsos positivos/negativos é minimizado para a predição de novos dados (ZORETTI; HARYNULK, 2011).

A técnica de classificação PLS-DA combina as propriedades da Regressão por Mínimos Quadrados Parciais (PLS) com a habilidade de classificação da Análise Discriminante (DA). O método dos mínimos quadrados parciais (PLS) proposto por Wold; Sjostrom, (2001), é o método mais usado em calibração multivariada e permite a relação entre informações espectrais (matriz \mathbf{X}) e as informações das concentrações (matriz \mathbf{Y}). No modelo PLS a decomposição da matriz \mathbf{X} é feita ponderando informações dadas em \mathbf{Y} , de forma que se obtenha a máxima covariância entre \mathbf{X} e \mathbf{Y} definindo assim as Variáveis Latentes (VL) do modelo. Na modelagem por mínimos quadrados parciais as matrizes \mathbf{X} e \mathbf{Y} são decompostas simultaneamente em matrizes menores (BALLABIO; CONSONNI, 2013)

2.11. Uso da espectroscopia NIR em sementes na diferenciação de genótipos

O sucesso de qualquer empreendimento agrícola produtivo é iniciado com a escolha do material que será implantado, na certeza que este, possua características genéticas estáveis e que a sua pureza varietal seja mantida em todo o processo de produção até a comercialização da semente ou propágulo.

A manutenção da pureza varietal se torna fundamental ao passo que carrega considerável complexidade, haja visto que, em algumas espécies as diferentes cultivares podem apresentar características físicas muito semelhantes entre si como tamanho, formato, peso, aparência, entre outras características o que dificulta e na maioria dos casos inviabiliza tecnicamente a separação das sementes de cultivares distintas no processo de colheita, beneficiamento bem como no processo de análise em laboratório.

No caso da mamona, para que a identificação de possíveis misturas varietais podem ser utilizadas as técnicas de avaliação morfológica na qual são avaliados aspectos como cor, cerosidade, tipo de folha, entre outros, o uso de marcadores moleculares e mais recentemente a espectroscopia do infravermelho próximo. Vale salientar que a avaliação por caracteres morfológicos geralmente é realizada após o trigésimo dia após a emergência das plântulas, uma

vez que, antes desse período, as diversas variedades podem estar sofrendo a diferenciação celular (MILANI *et al.*, 2009; SANTOS *et al.*, 2014).

A utilização de estratégias analíticas empregando técnicas instrumentais e análise multivariada tem permitido superar os desafios decorrentes da baixa eficiência analítica em produtos agrícolas e agroindustriais (AGELET, 2014).

Várias são as técnicas instrumentais de análise que buscam a classificação de sementes. As mais estudadas recentemente envolvem os métodos baseados na espectroscopia na região do infravermelho próximo (NIR) que estão sendo os mais difundidos e estudados (LI; HE; WU, 2008; LEE; CHOUNG, 2011; VITALE *et al.*, 2013).

Com o intuito de diferenciar genótipos de girassol quanto aos teores de ácido oleico e linoleico, Grunvald *et al.*, (2009) desenvolveram uma calibração para a discriminação dos referidos ácidos e esta se mostrou satisfatória para a seleção de genótipos de girassol com alto teor de ácido graxo oléico, além de ser mais rápida e não destrutiva à semente.

Lee; Choung, (2011) classificaram sementes de soja geneticamente modificadas (GM) e não GM utilizando a espectroscopia NIR em conjunto com técnicas de Análise de Componentes Principais (PCA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). Vitale *et al.*, (2013), utilizaram o NIR com a Modelagem Independente Flexível por Analogia de Classe (SIMCA) e PLS-DA para a identificação da origem das sementes de pistache (*Pistacia vera L.*)

Magrini *et al.*, (2013) avaliando o potencial da técnica de espectroscopia de refletância no infravermelho próximo NIR em caracterizar diferentes genótipos de gramíneas em duas épocas do ano concluiu que a técnica NIR é uma poderosa ferramenta a ser utilizada em programas de melhoramento de espécies vegetais e com esta é possível diferenciar os acessos como, também, detectar a interação genótipo por ambiente.

Recentemente, Santos *et al.*, (2014), aplicaram o SIMCA e o PLS-DA aos dados obtidos pela espectroscopia de reflectância NIR para a classificação de sementes de mamona com respeito ao tipo de cultivar (BRS Nordestina e BRS Paraguaçu). De acordo com os autores, os melhores resultados foram obtidos com o PLS-DA, alcançando uma taxa de classificação correta de 100 % para o conjunto de Teste. Apesar dos bons resultados obtidos por essa metodologia, seu desenvolvimento exigiu o uso de um equipamento de alto custo.

A utilização de métodos baseados em reflectância difusa na região do visível e de imagens digitais envolvem equipamentos mais simples e acessíveis a laboratórios de rotina (SANTOS *et al.*, 2014; VILAR *et al.*, 2014)

Em especial para produtos agrícolas, a região do visível com quimiometria tem sido explorada em algumas aplicações: Li; He; Wu, (2008), discriminaram sementes de arroz de tempos diferentes de colheita utilizando espectroscopia Vis-Nir, empregando a análise não supervisionada (PCA) e supervisionada (ANN - Rede Neural Artificial). Xie; Ying; Ying *et al.* (2007) usaram a espectroscopia Vis-Nir para classificação de tomates transgênicos e não transgênicos empregando PCA e PLS-DA. Li; Yang (2014), classificaram méis utilizando espectroscopia Vis-Nir aplicando PCA e modelos de classificação PC - LDA e PC - SVM.

3. REFERÊNCIAS BIBLIOGRÁFICAS

ABDULLATEEF, R. A.; OSMAN, M. Effects of Visible Light Wavelengths on Seed Germinability in *Stevia Rebaudiana* Bertoni, **International Journal of Biology**, v. 3, p. 83, 2011.

AGELET, L. E.; JR, C. R. H. Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta*, 121: 288, 2014.

ALCÁCER, L. Textos de apoio a Química-Física. Determinação da Estrutura Molecular. Métodos Espectroscópios. **AEIST**, (2007).

ALMEIDA, F. **Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na Caracterização de Farinhas para Alimentação Pueril**. Tese de Doutorado, Instituto Superior Técnico, 2009.

ALVES, J. C. L.; POPPI, R. J. "Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM)." *Talanta* 104 (0): 155-161. 2013.

ARAÚJO, A. M. **Monitoramento de processos em tempo real via espectrofotometria no infravermelho próximo**. Dissertação de mestrado, UNIVERSIDADE TIRADENTES, 2007.

ASSOCIATION OF OFFICIAL SEED ANALYST – AOSA. **Cultivar purity testing**. Lansing, 1991. 371p.

ASSOCIATION OF OFFICIAL SEED ANALYSTS – AOSA. Seed vigor testing handbook. Lansing, 1983. 88 p. (Contribution, 32).

AZEVEDO, D.M.P; LIMA, E.F. **O agronegócio da mamona no Brasil**. Brasília, DF: Embrapa Informação Tecnológica, 2001. 350 p.

AZEVEDO, D. M. P. Manejo Cultural. In: AZEVEDO, D. M. P. de; LIMA, E. F. (Eds). **O agronegócio da Mamona no Brasil**. Brasília: Embrapa Comunicação para Transferência de Tecnologia, cap. 10, p.223-253. 2007.

BALABIN, R. M., SAFIEVA, R. Z. Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. **Fuel**, v. 87, p. 1096, 2008.

BALLABIO, D., CONSONNI, V. Classification tools in chemistry. Part 1: linear models PLS-DA. **Analitycal Methods**. v. 5, p. 3, 2013.

BARBOSA, L.C. de A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos**. Viçosa: UFV, 2008.

BARBOZA, F. D. Desenvolvimento e aplicação de métodos de transferência de calibração em espectroscopia no infravermelho próximo de ondas curtas – Campinas-SP. Dissertação de mestrado, 2003.

BARKER, M.; RAYENS, W. Partial least squares for discrimination. *Journal of chemometrics*, 2003. 17(3). 166-173.

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. 25 Anos de quimiometria no Brasil. **Química Nova**, v. 29, p. 140, 2006

BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. **Chemometrics A Practical Guide**. New York: John Wiley & Sons, 1998.

BELTRÃO, N. E. M.; AZEVEDO, D. M. P. Fitologia. In: AZEVEDO, D. M. P. de; BELTRÃO, N. E. de M. (Eds). **O agronegócio da mamona no Brasil**. 2. ed. Brasília: Embrapa Comunicação para Transferência de Tecnologia, cap. 5, p.119-137. 2007.

BINOTTI, F. F. S.; HAGA, K. I.; CARDOSO, E. D.; ALVES, C. Z.; SÁ, M. E.; ARF, O. Efeito do período de envelhecimento acelerado no teste de condutividade elétrica e na qualidade fisiológica de sementes de feijão *Acta Scientiarum*. Agronomy, v.30,n.2,p.247-254,2008.

BLANCO M.; COELLO J.; ITURRIAGA, H.; MASPOCH S.; DE LA PEZUELA, C. Near-infrared spectroscopy in the pharmaceutical industry. **Analyst**. 1998; p.123-135 R–150R. doi: 10.1039/a802531b.

BONOW, S.; PINHO, E.V. de R. von; SOARES, A.A. Caracterização morfológica de cultivares de arroz visando a certificação da pureza varietal. **Ciência e Agrotecnologia**, Lavras, v.31, n.3, p.619-627, maio/jun. 2007.

BORÉM, A.; CAIXETA, E. T. Marcadores moleculares. 2. ed. Viçosa, MG: UFV, 2009. 532 p.

BOTELHO, B. G.; MENDES, B. A. P.; SENA, M. M. Implementação de um método robusto para o controle fiscal de umidade em queijo minas artesanal. Abordagem metrológica multivariada. **Química Nova**, vol.36 nº.9, p. 1416-1422, 2013.

BRANDEN, K. V.; HUBERT, M. Robust classification in high dimensions based on the SIMCA Method. **Chemometrics and Intelligent Laboratory Systems**, v. 79, p.10, 2005.

BRASIL. **Ministério de Agricultura, Pecuária e Abastecimento. Instrução Normativa no 52, de 7 de novembro de 2011.** *Ministério da Agricultura, Pecuária e Abastecimento*. Brasília: [s.n.], 2011

BREITKREITZ, M. C.; RAIMUNDO, I. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; DANTAS, H. A.; JOSE, G. E.; ARAUJO, M. C. U. Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. **Analyst**. 128: 1204, 2003.

BRERETON, R. **Chemometrics for Pattern Recognition**. John Wiley & Sons: Chichester, 2007.

BROWN, S.D. Chemical systems under indirect observation: Latent properties and chemometrics. **Applied Spectroscopy**, v. 49, p. 14-31, 1995.

BRUNS, R. R.; FAIGLE, J. F. G. Quimiometria. **Química Nova**, v. 8, p.84, 1985.

BUENO, A. F. **Desenvolvimento de um analisador de processo por espectroscopia no infravermelho próximo (NIR) para precisão de propriedades de derivados de petróleo**. 2011. 264 f. Tese (Doutorado Química) - Universidade Estadual de Campinas, Campinas, 2011.
CAMPOS, A. C. M. **Carvão de Eucalyptus: Efeito dos parâmetros da pirólise sobre a madeira e seus componentes químicos e predição da qualidade pela espectroscopia NIR**. Dissertação (Mestrado em Ciência e Tecnologia da Madeira) Universidade Federal de Lavras, Lavras, 118p. 2008.

CARNEIRO, M. E. **Classificação de lâminas de madeira de pinus spp por espectroscopia óptica**. 2008, 97 f. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2008.

CARVALHO, N.M.; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. 3ª ed. Campinas, SP: Fundação Cargil, 1988. 424p.

CASALE, M.; ZUNIN, P.; COSULICH, M. E.; PISTACRINO, E.; PEREGO, P.; LANTERI, S. Characterisation of table olive cultivar by NIR spectroscopy. **Food Chemistry**, v.122, p. 1261-1265, 2010.

CASTRO, C. C. A influência da incerteza no desenvolvimento de pesquisas com organismos geneticamente modificados no Brasil. 2006. 252 p. Tese (Doutorado em Agronegócios) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T., Utilização de filtro de transformada de fourier para a minimização de ruídos em sinais analíticos. **Química Nova**, v. 23, p.690, 2000.

CHAGAS, I. P. **Desenvolvimento de um Fotômetro Portátil NIR Para Determinação do Teor de Água no Álcool Combustível e do Teor de Etanol na Gasolina.** 2006. 151 f. Tese (Doutorado em Química) - Universidade Estadual de Campinas, Campinas, 2006.

CHEDIAK, G.L.C.; BRONDANI R. P. V.; DEL PELOSO, M. J.; MELO, L. C.; BRONDANI, C. **Análise de pureza genética de sementes de feijoeiro comum utilizando marcadores microsatélites em sistema de genotipagem multiplex.** Santo Antônio de Goiás, Embrapa Arroz e Feijão. Boletim de Pesquisa e Desenvolvimento 28, ISSN 1678-9601; 28, 20 p. 2007.

CHOER, E.; BARBIERI, R. L.; CASTRO, C. M. **Qualidade genética de sementes de feijão utilizadas em regiões produtoras do Estado do Rio Grande do Sul.** Pelotas: EMBRAPA Clima Temperado, (Boletim de Pesquisa e Desenvolvimento, 3). 18 p. 2002.

COATES, J. Interpretation of Infrared Spectra, A Practical Approach. In Meyers, R., (Ed.). *Encyclopedia of Analytical Chemistry*. Chichester John Wiley & Sons Ltda. 2000.

COSTA FILHO, P. A. DA. **Estudo comparativo entre técnicas de inteligência artificial e modelos lineares em determinações quantitativas no infravermelho próximo.** Tese (Doutorado) - Instituto de Química, Universidade Estadual de Campinas, Campinas - SP. 2003

COSTA, M.N.; PEREIRA, W. E.; BRUNO, R. L. A.; FREIRE, E. C.; NOBREGA, M. B. M.; MILANI, M.; OLIVEIRA, A. P., Genetic divergence on castor bean accesses and cultivars through multivariate analysis. **Pesquisa Agropecuária Brasileira**, v. 41, 1617-1622, 2006.

DASZYKOWSKI, M.; KACZMAREK, K.; VANDER-HEYDEN, Y.; WALCZAK, B., Robust statistics in data analysis — A review Basic concepts. **Chemometrics and Intelligent Laboratory Systems**, v. 85, p. 203-219, 2007.

DERDE, M.P.; MASSART, D.L. Comparison of the Performance of the Class Modelling Techniques UNEQ, SIMCA, and PRIMA. **Chemometrics and Intelligent Laboratory Systems**, v.4, p. 65-93, 1988.

DINIZ, P. H. G. D.; DANTAS, H. V.; MELO, K. D. T.; BARBOSA, M. F.; HARDING, D. P.; NASCIMENTO, E. C. L.; PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C. U.; Using a simple digital camera and SPA-LDA modeling to screen teas. **Analytical Methods** v. 4, p. 2648–2652. 2012 doi:[10.1039/c2ay25481f](https://doi.org/10.1039/c2ay25481f)

EWING, G.W. **Métodos instrumentais de análise química**. Tradução: ALBANESE, A. G.; CAMPOS, J. T. de S. São Paulo: Blucher, 2011.

FERNANDES, D. D. S.; GOMES, A. A.; COSTA, G. B.; SILVA, G. W. B.; VÉRAS, G. Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection. **Talanta** v. 87, p. 30-34, 2011.

FERREIRA, C. A.; PINHO, E.V. de R. von; PAIVA, P. D. O.; PEREIRA, G. S.; DINIZ, R. P., Identificação de cultivares e certificação da pureza genética de gladiolo por meio de marcadores morfológicos. **Ciência e Agrotecnologia**, Lavras, v. 35, n. 4, p. 692-700, jul./ago. 2011.

FERREIRA, M. A. J. da F. **Utilização das técnicas de marcadores moleculares na genética de populações, na genética quantitativa e no melhoramento de plantas**. Boa Vista: Embrapa Roraima,. Documento, 1. 2003. 63p.

FERREIRA, M. M. C. Multivariate QSAR. **Journal of the Brazilian Chemical Society**, v. 13, p. 742-753, 2002.

FERREIRA, M. M. C., MONTANARI, C. A., GAUDIO, A. C. Seleção de Variáveis em QSAR. **Química Nova**, v. 25 (3), p. 439-448, 2002.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p.724-731, 1999.

FIGUEIREDO NETO, A.; ALMEIDA, F. A.C.; GOUVEIA, J.P.G.; NÓBREGA, M.B.M.; CARNEIRO, R.M.; PEDROZA, J.P. Divergência genética em acessos de mamona (*Ricinus communis* L.) baseada nas características das sementes. *Revista de Biologia e Ciências da Terra*, v.4, n.002, 2004.

FLATEN, G. R.; GRUNG, B.; KVALHEIM, O. M. A method for validation of reference sets in SIMCA modeling. **Chemometrics and Intelligent Laboratory Systems**, v.72, p.101-109, 2004.

FLUMIGNAN, D. L. **Caracterização da qualidade e precisão dos parâmetros físico-químicos de gasolinas comerciais brasileiras através da aplicação de métodos quimiométricos em perfis (fingerprintings) espectroscópicos de ressonância magnética nuclear**. 2010. 225f. Tese (Doutorado em Química). Universidade Estadual Paulista, Araraquara, 2010.

FREIRE, E.C.; LIMA, E.F.; ANDRADE, F.P. Melhoramento genético. In: AZEVEDO, D.M.P. de; LIMA, E.F. (Ed.). *O agronegócio da mamona no Brasil*. Brasília: Embrapa Informação Tecnológica, 2001. p.229-256.

GALVÃO, R. K. H.; ARAUJO, M. C. U.; SALDANHA, T. C. B.; VISANI, V.; PIMENTEL, M. F. Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de Fourier e Wavelet. **Química Nova**, v. 24, p.874-884, 2001.

GAMBARRA-NETO, F.F. **Classificação de óleos vegetais utilizando voltametria de onda quadrada e métodos quimiométricos**. João Pessoa, Programa de Pós-Graduação em Química, UFPB,. Dissertação de Mestrado, 98p. 2008

GELADI, P. "Chemometrics in spectroscopy. Part 1. Classical chemometrics." *Spectrochimica Acta Part B: Atomic Spectroscopy*, v. 58(5), p. 767-782, 2003.

GELADI, P.; KOWALSKI, B. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.

GEMPERLINE, P. **Practical guide to chemometrics**. 2nd edição. Boca Raton, FL, 2006.

GRUNVALD, A. K.; LEITE, R. S.; TERRA, I. M.; CARVALHO, C. G. P. de; MANDARINO, J. M. G.; ANDRADE, C. A. B.; OLIVEIRA, A. C. B. de; RAMOS, N. P.; AMABILE, R. F.; GODINHO, V. de P. C.; CARVALHO, H. W. L. de; OLIVEIRA, I. R. de; AZEVEDO, N. A.;

GONCALVES, S. L. **Curva de calibração para predição do teor de óleo pela análise do espectrômetro de infravermelho próximo (NIR) em ensaios de competição de cultivares de girassol no Brasil.** In: REUNIÃO NACIONAL DE PESQUISA DO GIRASSOL, 18.; SIMPÓSIO NACIONAL SOBRE A CULTURA DO GIRASSOL, 6., 2009, Pelotas. Anais... Pelotas: Embrapa Clima Temperado, p. 51-58. 2009.

HEIN, P. R. G.; CAMPOS, A. C. M.; LIMA, J. T.; TRUGILHO, P. F.; CHAIX, G. Estimativa da resistencia e da elasticidade a compressao paralela as fibras da madeira de *Eucalyptus grandis* e *E. urophylla* usando a espectroscopia no infravermelho proximo. **Scientia Forestales**, Piracicaba. V37, n82, p.119-129, jun. 2009.

HIGSON, S. P. J. **Química Analítica.** São Paulo: McGraw-Hill, 2009.

HOMMA, S., EDA, H., OGASAWARA, S., KAGAYA, A. **Near-Infrared Estimation Of O2 Supply And Consumption.** In Forearm Muscles Working At Varying Intensity, J. v. 80(04), p. 1279-1284, 1996.

HONORATO, F. A. **Previsão das propriedades das gasolinas do Nordeste empregando espectroscopia NIR/MID e transferência de calibração.** 2006. 106 f. Tese (Doutorado em Química) – Universidade Federal de Pernambuco, Recife, 2006.

INTERNATIONAL SEED TESTING ASSOCIATION – ISTA. **Rules for seed testing.** Zürich, 1996. 44p.

KAMAL-ELDIN, A.; ANDERSSON, R. A Multivariate Study of the Correlation Between Tocopherol Content and Fatty Acid Composition in Vegetable Oils. **Journal of the American Oil Chemists' Society**, v. 74, p. 375-380, 1997.

KENNARD, R. W.; STONE, L. A., Computer-aided design of experiments. **Technometrics**. v.11, p. 137-148, 1969.

KILLNER, M. H. M.; ROHWEDDER, J. J. R.; PASQUINI, C. A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction. **Fuel**, v. 90, n. 11, p. 3268–3273, nov. 2011.

KIM, J., MOWAT, A., POOLE, P., KASABOV, N., **Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra**. Chemom. Intell. Lab. Syst. 51, 201–216, 2000.

KOWALSKI, B. R. **Chemometrics, Mathematics and Statistics in Chemistry**. Consenza, Springer, 1983.

KUMAR, N.; BANSAL, A.; SARMA, G. S.; RAWAL, R. K., Chemometrics tools used in analytical chemistry: An overview. **Talanta**, v. 123, p.186-199, 2014.

LEE, J. H.; CHOUNG, M. G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near infrared reflectance spectroscopy. **Food Chemistry**, v. 126, p. 368-373, 2011.

LI, X.; HE, Y.; WU, C. Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. **Journal of Stored Products Research**, v. 44, p.264-268, 2008.

Li, Y.; Yang, H. **Honey Discrimination Using Visible and Near-Infrared Spectroscopy**, ISRN Spectroscopy, vol. 2012, Article ID 487040, 4 pages, 2012. doi:10.5402/2012/487040

LIMA, K. M. G. et al. Sensores ópticos com detecção no infravermelho próximo e médio. **Química Nova**, v. 32, p.1635-1643, 2009

LIRA, L. F. B. **Desenvolvimento de métodos analíticos para monitoramento da qualidade do biodiesel e suas misturas**. 2010. 145 f. Tese (Doutorado em Química) – Universidade Federal de Pernambuco, Recife, 2010.

LUZ, E. R.; SILVA, M. I. P.; TÉLLEZ, C. A. S. PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO Departamento de Química. **Predição de propriedades de gasolinas**

usando espectroscopia FTIR e regressão por mínimos quadrados parciais. 2003. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Química, Rio de Janeiro, 2003

MAGRINI, V.; DIAS, M.; COSTA, A. C. da S.; SOUZA, G. B. de; MATTA, F. de P.; VIGNA, B. B. Z.; FAVERO, A. P. **A espectroscopia de refletância no infravermelho próximo e a interação genótipo x ambiente.** In: ENCONTRO LATINO AMERICANO DE INICIAÇÃO CIENTÍFICA, 17.; ENCONTRO LATINO AMERICANO DE PÓS GRADUAÇÃO, 13.; ENCONTRO LATINO AMERICANO DE INICIAÇÃO CIENTÍFICA JÚNIOR, 7.; ENCONTRO DE INICIAÇÃO À DOCÊNCIA, 3., 2013., São José dos Campos. Anais... São José dos Campos: UNIVAP, 2013.

MARCOS FILHO, J. Testes de vigor: importância e utilização. In: KRZYZANOWSKI, F.C.; VIEIRA, R.D.; FRANÇA NETO, J.B. (Ed.). **Vigor de sementes: conceitos e testes.** ABRATES, cap.1, p.1-21. 1999.

MARTENS, H., NAES T. **Multivariate Calibration.** John Wiley: New York, 1989.

MARTINS, S. M.; MUNIZ, D. K. Interação entre Filmes de PVC e Chorume Enriquecido com Meio Mineral. **Arq. Inst. Biol.**, 69(3): p.103-107, 2002.

MASSART, D. L.; VANDEGINSTE, L. M. C.; BUYDENS, S. J.; LEWI. P.; SMEYERS-VERBEKE, J., **Journal Handbook of Chemometrics and Qualimetrics: Parte B,** Amsterdam: Elsevier, 1997.

MILAGRES, F. R. **Avaliação da madeira de híbridos de Eucalyptus globulus com E. grandis e E. urophylla, para a produção de celulose, utilizando a espectroscopia NIR.** Dissertação (Mestrado em Ciências Florestais) Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo. Piracicaba, 2009.

MILANI, M.; MIGUEL JÚNIOR, S. R.; SOUSA, R. de L. **Subespécies de mamona.** Campina Grande: Embrapa Algodão, 23p. Documento, 230. 2009.

MOITA NETO, J.M.; MOITA, G.C. Uma introdução à análise exploratória de dados multivariados, **Química Nova**, v. 21 (4), p. 467-469, 1998.

MORALES, Á. D.; OSORIO, A. C.; FERNANDEZ, R. F.; DENNES, E. L. Desarrollo de un modelo SIMCA para la clasificación de kerosinas mediante el empleo de la espectroscopía infrarroja. **Química Nova**, v. 31, n. 6, p. 1573-1576, 2008.

MOREIRA, J. A. N.; LIMA, E. F.; FARIAS, F. J. C.; AZEVEDO, D. M. P. **Melhoramento da mamoneira (*Ricinus communis* L.)**. Campina Grande: Embrapa Algodão, 29p. Documento, 44. 1996.

MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F.; BRAGAGNOLO, N.; FERREIRA, M. C. Determinação de proteína em café cru por espectroscopia NIR e Regressão PLS. Campinas: **Ciência e Tecnologia de Alimentos**, v. 25, n.1, p. 25-31, 2005.

MOSHKIN, V. A. **Castor**. New Delhi: Amerind, 1986. 315 p.

MOSHKIN, V.A.; PERESTOVA, T.A. Morphology and anatomy. In: MOSHKIN, V.A. (ed.). **Castor**. New Delhi: Amerind, 1986. p. 28-33.

NAES, T.; ISAKSSON, T.; FEARN T.; DAVIES T., **A User-Friendly Guide to Multivariate Calibration and Classification**. Chichester, UK: NIR Publications, 2002.

NETO JÚNIOR. G. V., **Um espectrômetro microcontrolado baseado em LED branco como fonte de radiação e mídia de CD como grade de difração**. 116f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba. 2008.

NISGOSKI, S. **Espectroscopia no infravermelho próximo no estudo de características da madeira e papel de *Pinus taeda* L.** Tese (Doutorado em Ciência Florestais) Universidade Federal do Paraná, Curitiba, 2005.

NUNES, P. G. A. **Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às Espectrometrias UV-VIS E NIR**. 2008. 121 f. Tese (Doutorado em Química) – Universidade Federal da Paraíba, João Pessoa, 2008.

OTTO, M. **Chemometrics: statistics and computer application in analytical chemistry**, New York: Wiley-VCH, 1999.

PANTOJA, P. A. **Caracterização do petróleo por técnicas espectroscópicas e sua utilização na avaliação e desenvolvimento de métodos de inferência em processo de refino**. 162f. Tese (Doutorado) - Departamento de Engenharia Química, Escola Politécnica da Universidade de São Paulo, São Paulo - SP. 2010.

PARIZZI, F. C.; FARONI, L. R. D.; PEARSON, T. C.; DHINGRA, O. D. **Espectroscopia de Infravermelho Próximo na Detecção de Defeitos e Fungos em Grãos de Café**. In: IV Seminário Nacional de Armazenagem, Uberlândia, 2005.

PASQUINI, C. **Espectroscopia no Infravermelho Próximo (NIR)**. Salvador, UFBA, 2002.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, p. 198-219, 2003.

PAVIA, D.L.; LAMPMAN, G.M., and KRIZ; G.S., **Introduction to Spectroscopy: A Guide for Students of Organic Chemistry**. Third ed. 2001: Thomson Learning.

PEREIRA, A. F. C.; PONTES, M. J. C. GAMBARRA NETO, F. F.; SANTOS, S. R. B.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection. **Food Research International**. v. 41, p. 341-348, 2008.

PONTES, M. J. C. de. **Algoritmo das projeções sucessivas para a seleção de variáveis espectrais em problemas de classificação**. 2009. 144 f. Tese (Doutorado em Química)– Universidade Federal da Paraíba, João Pessoa, 2009.

PONTES, M. J. C.; ROCHA, A. M. J.; PIMENTEL, M. F.; PEREIRA, C. F. Determining the quality of insulating oils using near infrared spectroscopy and wavelength selection. **Microchemical journal**, v. 98, p. 254-259, 2011.

PONTES, M.J.C.; GALVÃO, R.K.H.; ARAÚJO, M.C.U.; MOREIRA, P.N.T.; NETO, O.D.P.; JOSÉ, G.E.; SALDANHA, T.C.B. **The successive projections algorithm for spectral variable selection in classification problems.** Chemometrics and Intelligent Laboratory Systems, v. 78, p.11-18, 2005.

POVIA, G. S. **Determinação dos Parâmetros de Qualidade de Detergentes em Pó Utilizando Espectroscopia no Infravermelho Próximo.** 2007. 84 f. Dissertação (Mestrado em Química) - Universidade Estadual de Campinas, Campinas, 2007.

RAMOS, N. P. **Determinação da pureza varietal em lotes de sementes de milho através de marcadores morfológicos e microssatélites.** 2004. 104 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiróz”, Piracicaba, 2004.

REICH, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. **Advanced Drug Delivery Reviews**, v. 57, p. 1109-1143, 2005.

RODRIGUES, C. R.; Espectrofotometria por infravermelho. **Escola Técnica do Brasil Santos Dumont.** Jundiaí, 2012. Disponível em: https://www.academia.edu/5328536/ESPECTROFOTOMETRIA_POR_INFRAVERMELHO

SABIN, J. G.; FERRÃO, M. F.; FURTADO, J. C. Análise multivariada aplicada na identificação de fármacos antidepressivos. Parte II: Análise por componentes principais (PCA) e o método de classificação SIMCA. **Revista Brasileira de Ciências Farmacêuticas**, v. 40, p. 387-396, 2004.

SANCHES, Fátima Aparecida Castriani. **Desenvolvimento de uma metodologia para o controle de qualidade de medicamento injetável, sem violação da ampola, usando espectroscopia NIR e técnicas quimiométricas.** 2009. 72 f. Dissertação – (Programa de Pós-Graduação em Química–Mestrado) - Universidade Federal da Paraíba, João Pessoa, 2009.

SANTOS, A.F.; LIMA, E.L.; PINTO, J.C. Control and Design of Average Particle Size in Styrene Suspension Polymerization Using NIRS. **Journal of Applied Polymer Science**, v. 77, p.453-462, 2000.

SANTOS, M. B. H., GOMES, A. A., VILAR, W. T. S., ALMEIDA, P. B. A., MILANI, M., NÓBREGA, M. B. M., MEDEIROS, E. P., GALVÃO, R. K. H., ARAÚJO, M. C. U. Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. **Journal Brazilian Chemistry Society**, v. 25, p. 969-974, 2014.

SANTOS, P. **Aplicação da espectroscopia de infravermelhos próximo na avaliação de alguns atributos de qualidade em pêra “rocha” e maçã “gala”**. Tese de Doutorado, Instituto Superior de Agronomia, 2009.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least-squares procedures. **Analytical Chemistry**, v. 36 p. 1627-1639, 1964.

SAVY FILHO, A. **Mamona Tecnologia Agrícola**. Campinas: EMOPI, 2005. 105 p.

SCAFI, S. H. F. **Espectroscopia no Infravermelho Próximo para identificação de medicamentos falsificados**. Dissertação (Mestrado em Química) - Universidade Estadual de Campinas, 139 f, 2000.

SCAFI, S. H. F. **Sistema de Monitoramento em Tempo Real de Destilações de Petróleo e Derivados Empregando a Espectroscopia no Infravermelho Próximo**. Tese (Doutorado em Química) - Universidade Estadual de Campinas, Campinas, 214 f, 2005.

SENA, M. M.; FRIGUETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v. 23, p. 441-446, 2000.

SENA, M. M.; POPPI, R. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, v. 23, n. 4, p. 547-556, 2000.

SEVERINO, L. S.; AULD, D. L.; BALDANZI, M.; CÂNDIDO, J. D.; CHEN, G.; CROSBY, W.; TAN D.; XIAOHUA HE; LAKSHMAMMA, P.; LAVANYA, C.; MACHADO, O. L. T.; MIELKE, T.; MILANI, M.; MILLER, T. D.; MORRIS, J. B; MORSE, S. A.; NAVAS A. A.;

SOARES, D. J.; SOFIATTI, V.; WANG, M. L.; ZANOTTO, M. D.; ZIELER, H. A Review on the Challenges for Increased Production of Castor. **Agronomy Journal**, v. 104, p. 853-880, 2012.

SIESLER, H. *et al.* **Near-infrared spectroscopy: principles, instruments, applications.** Weinheim, Wiley-VCH, v. 3, p. 903-904, 2002.

SILVA, H., RODRIGUES, M., E EUSÉBIO, M. **Espectroscopia no infravermelho por transformada de Fourier (FTIR): fundamentos e aplicação na caracterização de materiais de construção.** Lisboa, LNEC, 2008.

SIMÕES, S. S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada.** 2008. 98 f. Tese (Doutorado em Química) – Universidade Federal da Paraíba, João Pessoa, 2008.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Princípios de Análise Instrumental.** 6a. ed. Porto Alegre: Bookman, 2009. p. 1056

SKOOG, D. A.; WEST, D. M.; HOLLER, F. J.; CROUCH, S. R. **Fundamentals of Analytical Chemistry.** 9. Belmont: Cengage Learning, 2014.

SOARES, S. F. C., GALVÃO, R. K. H., PONTES, M. J. C., ARAÚJO, M. C. U. A New Validation Criterion for Guiding the Selection of Variables by the Successive Projections Algorithm in Classification Problems. **Journal Brazilian Chemistry Society**, v. 25, p.176-182, 2014.

SOARES, S. F. C.; GOMES, A. A.; GALVÃO FILHO, A. R.; ARAUJO, M. C. U.; GALVÃO, R. K. H. The successive projections algorithm, **Trends in Analytical Chemistry**, v. 42, p.84-98, 2013.

SOUZA, A.M.; POPPI, R.J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Quim. Nova**, Vol. 35, p. 223-229, 2012.

STUMPE, B.; ENGEL, T.; STEINWEG, B.; MARCSCHNER, B., Application of PCA and SIMCA Statistical Analysis of FT-IR Spectra for the Classification and Identification of Different Slag Types with Environmental Origin. **Environmental Science Technology**, v.46, p. 3964-3972, 2012.

VERAS, G.; BRITO, A. L. B.; SILVA, A. C.; SILVA, P.; COSTA, G. B.; FÉLIX, L. C. N.; FERNANDES, D. D. S.; FONTES, M. M. Classificação de Biodiesel na região do visível. **Química Nova**, p.1-4, 2011.

VIDAL, M. S.; MILANI, M.; MENEZES, C.H. S. G.; BEZERRA, C. S., **Seleção de Marcadores do Tipo Rapd para Caracterização Genética *Ricinus communis* L.** Campina Grande: Embrapa Algodão, Documento, 90. 5p. 2005.

VILAR, W. T. S., Classificação individual de sementes de mamona usando espectroscopia de reflectância no visível, imagens digitais e análise multivariada, 2014. 103 f. **Dissertação** (Mestrado em Química). João Pessoa-PB, 2014.

VITALE, R.; BEVILACQUA, M.; BUCCI, R.; MAGRI, A. D.; MAGRI, A. L.; MARINI, F. A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 121, p. 90-99, 2013.

WATSON, C.A. Near Infrared Reflectance Spectrophotometric Analysis of Agricultural Products. **Analytical Chemistry**, v. 49, p. 835-840, 1977.

WEHRENS, R. **Chemometrics With R: Multivariate Data Analysis in the Natural Sciences and Life Sciences**, Springer, 2011.

WOLD, S. Pattern recognition by means of disjoint principal components models. **Pattern Recognition**, v.8, p.127-139, 1976.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal Component Analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 2, p. 37-52, 1987.

WORKMAN, J., **The Handbook of Organic Compounds**. v. 3. ACADEMIC PRESS. 2000.
XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J.W.; HOLMES, M.; HANPIN, M. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, p. 14–32, 2010.

XIE, L.J.; YING, Y.B.; YING, T.J. Combination and comparison of chemometrics methods for identification of transgenic tomatoes using visible and near-infrared diffuse transmittance technique. **Journal of Food Engineering**. 82: p. 395-401, 2007.

YU, H.; YANG, J. A direct LDA algorithm for high-dimensional data with application to face recognition. **Pattern Recognition**, 34: 2067, 2001.

YUAN, H.; CHU, X.; WANG, Y.; YANG, Y.; LU, W. The Development of an On-line NIR Analyzer and its Applications in Catalytic Reforming Process. **Research Institute of Petroleum Processing**, 2003.

ZORZETTI, B. M.; HARYNUK, J. J. Using GC x GC-FID profiles to estimate the age of weathered gasoline samples. *Analytical and bioanalytical chemistry*. 2011. 401(8). 2423-2431.

CAPITULO I
MARCADORES
ESPECTRAIS SWNIR EM
CULTIVARES DE
MAMONA

FERREIRA, D.F. **Marcadores espectrais na região do SWNIR para seleção de sementes de mamona (*Ricinus communis* L.)**. 2015. 134f. Tese (Doutorado em Agronomia), Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba.

RESUMO

A mamona (*Ricinus comunnis* L.) é uma planta que apresenta alta variabilidade fenotípica, produtiva e de composição química. Muitas destas características estão relacionadas a vários genes que constituem o indivíduo e por sua vez podem ser bastante influenciados por estresses bióticos e abióticos, impossibilitando em muitos casos a distinção visual entre diferentes materiais. A busca dessa caracterização de forma eficiente e rápida tem despertado interesse as mais diversas áreas do conhecimento, e nesse sentido surge a espectroscopia no infravermelho. O objetivo deste estudo é estabelecer marcadores espectrais de identificação de genótipos de mamona baseado em cultivares já estáveis. Foram utilizados cinco genótipos de mamona no estudo, sendo utilizadas 50 sementes de alta qualidade genética cedidas pela Embrapa Algodão. De cada material, foram pesadas e tomadas suas dimensões e, em seguida acondicionadas a uma temperatura de 21°C e umidade relativa de 70% para realização das avaliações. As medidas de reflectância difusa foram obtidas em um espectrômetro VIS-NIR modelo XDS Rapid ContentTM Analyser (Foss Analytical, Hogans, Sweden). Os espectros de reflectância foram obtidos diretamente sem nenhum tratamento químico das sementes. Devido as similaridades observadas nos materiais analisados, foi feita a separação em dois grupos de trabalho sendo (A) BRS Nordestina e BRS Paraguaçu e (B) BRS Energia, CNPAM 2009-7 e BRS Gabriela. Os espectros de reflectância foram registrados na faixa de 750 a 1100nm que corresponde a região do infravermelho de ondas curtas SWNIR. Para os modelos de classificação foram utilizadas 50 sementes de cada material, sendo utilizadas para o conjunto de teste por validação cruzada, 36 sementes de cada. As leituras foram efetuadas com as sementes sempre na mesma posição. Os espectros foram pré-processados utilizando o algoritmo Savitzky-Golay com janela de 11 pontos e primeira derivada para correção da linha de base. Mediante a PCA (Principal Component Analysis) foi observado que a faixa em estudo apresentou resultados que permitiram a distinção entre os genótipos. O modelo SIMCA (Soft Independent Modelling of Class Analogy) demonstrou resultados promissores para classificação nos níveis de significância de 1,5 e 10%. O SPA-LDA (Sucessive Projections Algorithm-Linear Discriminant Analysis) foi eficiente classificando corretamente todas as amostras. Os métodos em estudo foram comprovadamente promissores para diferenciação fenotípica para genótipos de mamoneira de forma rápida, prática e não destrutiva.

Palavras-chave: Produção de sementes, *Ricinus comunnis* L., Espectroscopia SWNIR, Genótipos, SIMCA, SPA-LDA

FERREIRA, D.F. **Spectral markers in SWNIR region for selection seeds of castor beans (*Ricinus communis* L.)**. 2015. 134f. Thesis (Doctorate in Agronomy Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba, Areia, Paraíba.

ABSTRACT

The Castor bean (*Ricinus comunnis* L.) is a plant that has high phenotypic variability, productive and of chemical composition. Many of these features are related to several genes that constitute the individual and may be quite influenced by biotic and abiotic stresses, making it impossible in many cases the visual distinction between different materials. The search of this characterization of efficiently and quickly has awakened interest in the various areas of knowledge, and in this sense the infrared spectroscopy appears. The objective of this study is to establish spectral markers for identification of genotypes based on castor varieties already stable. Five genotypes of castor were used in the study, have been used 50 seeds of high genetics quality provided by Embrapa Cotton. From each material, were weighed and taken its your dimensions and then packed at a temperature of 21° C and 70% relative humidity for carrying out the analyzes. Diffuse reflectance measurements were obtained in a spectrometer VIS-NIR ContentTM model XDS Rapid Analyzer (Foss Analytical, Hogans, Sweden). The reflectance spectra were obtained directly without any chemical treatment of the seeds. Because of the similarities observed in the materials analysed, these were split into two working groups being (A) BRS Nordestina and BRS Paraguaçu and (B) BRS Energy, CNPAM 2009 -7 and BRS Gabriela. The reflectance spectra were recorded in the range of 750 to 1100nm that matches the region infrared shortwave -SWNIR. For models of classification that were used 50 seeds of each material, being used for the test by cross-validation, 36 seeds each. The readings were made with the seeds always in the same position. The spectra were preprocessed using the algorithm Savitzky-Golay with window of 11 points and first derivative for baseline correction. Trough PCA (Principal Component Analysis) it was observed that the study range showed results that allowed the distinction between the genotypes. The SIMCA (Soft Independent Modeling of Class Analogy) model demonstrated promising results for classification at the significance levels of 1.5 and 10%. The SPA-LDA (Sucessive Projections Algorithm-Linear Discriminant Analysis) was efficient sorting correctly the all samples. The methods under study were demonstrably been proven to be promising for phenotypic differentiation castor bean genotypes in a fast, practical and non-destructively.

Keywords: Seed production, *Ricinus comunnis* L., Spectroscopy SWNIR, Genotypes, SIMCA, SPA-LDA

1. INTRODUÇÃO

A mamoneira é uma planta que apresenta alta variabilidade entre as cultivares, envolvendo a coloração do caule, das folhas e das sementes, a deiscência dos frutos, a presença ou não de cerosidade, a altura das plantas, o ciclo, a produtividade, e o teor de óleo e de ricina nas sementes. Alguns destes caracteres muitas vezes são controlados por vários genes e bastante influenciados pelo ambiente, desta forma, informações sobre a amplitude dos coeficientes de variabilidade e de herdabilidade nos permitem prognosticar o efeito da seleção e planejar os procedimentos do melhoramento o que torna imprescindível o conhecimento da natureza dos caracteres agrônômicos da cultura (MOSHKIN, 1986; SAVY FILHO, 2005; BELTRÃO; AZEVEDO, 2007).

As regras da International Seed Testing Association (ISTA, 1995), o manual da Association of Official Seed Analysis (AOSA, 1983) e as Regras para Análise de Sementes (BRASIL, 2009), recomendam a caracterização de espécies e cultivares por meio das características morfológicas de sementes, plântulas e plantas, sendo o método mais antigos e amplamente difundidos na avaliação de diferentes cultivares. Apresentam como vantagens a simplicidade, a rapidez e o baixo custo de análise (BRETTEING; WIDRLECHENER, 1995).

Apesar disto, nem sempre é possível identificar diferentes genótipos utilizando a inspeção visual por meio de seus descritores morfológicos pois, geralmente a identificação é realizada em plantas inteiras ou adultas com no mínimo 30 dias após a germinação, podendo estas, serem influenciadas pelo ambiente, ser alteradas pelo efeito de um determinado patógeno, e ainda influenciados por interações intra e inter-loci, resultando em dados pouco confiáveis, apresentando problemas na identificação de plantas com base genética muito estreita (PADILHA et al., 2002; PRIOLLI et al., 2002).

Para a identificação e classificação de diferentes genótipos também são empregados os marcadores moleculares, (VECCHIA et al., 1998; FERREIRA, 2003; VIDAL et al., 2005). Estes constituem metodologia mais precisa, uma vez que, identifica diretamente a base genética do genótipo, no entanto, são difíceis de serem implantadas em escala de rotina, destroem a semente, inviabilizando-as para futuros testes; demandam um maior período de tempo e necessitam de pessoal com alta qualificação técnica.

Em um programa de melhoramento, principalmente nas fases iniciais, a disponibilidade de sementes é pequena e usar métodos que destruam as sementes para caracterização pode não ser viável. Além do mais, para realizar estudos que demandam mais tempo podem representar

atraso no programa de melhoramento, haja visto, que em muitos casos as gerações seguintes precisam seguir um calendário agrícola específico.

Com o intuito de superar esses desafios, os métodos analíticos baseados no uso da espectrometria de reflectância no infravermelho próximo (NIR) e as técnicas quimiométricas que permitem dentre outras vantagens a não destruição das amostras e a rapidez nas avaliações podem contribuir para a solução destes desafios.

A espectroscopia NIR associada as diversas técnicas quimiométricas, vem ganhando destaque na utilização em várias matrizes, dentre elas: café (RIBEIRO; SALVA, 2011), leite (ROCHA, 2010), determinação do teor de carbono no solo (SEGNINI et al., 2014), textura do solo (FERRARESI et al., 2012), madeira (BASTOS et al., 2012; SILVA et al., 2014), carvão (SILVA et al., 2014), determinação de micotoxinas (HORN, 2013), farinha de mandioca (FOLHA, 2014), nitrogênio em folhas (SANTOS; SANTOS; KORNDÖRFER, 2012), biodiesel (VERAS et al., 2010; BALABIN; SAFIEVA, 2011; INSAUSTI et al., 2012), gasolina (BALABIN; SAFIEVA; LOMAKINAC, 2010); cigarros (MOREIRA et al., 2009), cerveja (EGIDIO et al., 2011; GHASEMI-VARNAMKHAHI et al., 2012), pêssegos (BETEMPS et al., 2014), azeite (SINELLI et al., 2010; GALTIER et al., 2011); vinhos (RIOVANTO et al., 2011), mel (CHEN et al., 2012) e na construção de biblioteca estelar (DINIZ; RIFFEL, 2011)

Mesmo com a eficiência comprovada da técnica nas mais diversas áreas do conhecimento tanto para amostras sólidas quanto para líquidas, a espectroscopia NIR ainda ocupa pouco espaço quando se trata do screening de germoplasmas baseados na avaliação da composição química em sementes intactas (OSBORNE, 2006; TALLADA et al., 2009).

Marquete (2014) analisando 18 amostras de quatro genótipos de café utilizando cultivadas em quatro locais distintos destacaram que mesmo os genótipos tendo a composição genética similar, os espectros NIRS obtidos na faixa de comprimento de onda de 1100 a 2498nm apresentaram informações importantes para a discriminação dos mesmos. O método PLS-DA, com pré-processamento de correção do espalhamento multiplicativo e a segunda derivada conjuntamente, foi capaz de classificar as amostras de café arábica tanto geograficamente quanto genotipicamente.

Jiang et al., (2007) avaliando 309 amostras oriundas de material de reprodução de milho de vários locais e diversos anos, desenvolveu estudo com a finalidade de estabelecer as calibrações de predição robustas para proteína bruta, amido bruto e gordura bruta. Após a aquisição dos espectros de reflectância original, estes, foram processadas por cinco métodos matemáticos: Normalização de vetores (NV), correção do sinal multiplicativo (MSC), primeira derivada (1º

Deriv.), Primeiro Deriv. + MSC e primeira Deriv. + NV. Equações de calibração foram desenvolvidas utilizando o Software Bruker Optics OPUS / QUANT 4.0 para Windows com base no Método (PLS). Amostras consideradas *outlier* foram excluídas automaticamente pelo software de calibração. Os resultados demonstraram que era possível rastrear e selecionar proteínas, amido e a concentração de óleo ao nível do núcleo de sementes intactas com NIRS. Também foi demonstrado que as calibrações foram confiáveis e o método robusto o suficiente para atender a demandas de seleção recorrente no melhoramento havendo estreita correlação.

Pesquisas desenvolvidas por Lee; Choung (2011) para avaliar o potencial da espectroscopia NIR na classificação de sementes de soja geneticamente modificadas (GM) e não-GM. Espectros NIR foram coletados a partir das sementes individuais em que cada semente foi colocada em um suporte permitindo que a radiação fosse refletida de um lado da semente. Todos os dados espectrais foram registrados como o logaritmo do inverso da reflectância ($\log 1 / R$) na região espectral de 400 a 2500 nm, com resolução de 2 nm e média de 32 varreduras. As técnicas quimiométricas utilizadas foram análise de componentes principal (PCA) e análise discriminante por mínimos quadrados parciais (PLS-DA). O modelo PLS-DA usando os dados pré-processados, obteve a melhor calibração e um acerto na classificação de 97%. De acordo com os autores, os resultados com a espectroscopia NIR em conjunto com técnicas quimiométricas, podem ser usados para identificar soja GM evitando, assim, análises demoradas, destrutivas e trabalhosas.

Vitale et al., (2013) estudaram o potencial da espectroscopia NIR acoplada a técnicas quimiométricas (SIMCA, PLS-DA) para verificar a origem de sementes de pistache (*Pistacia vera* L.). Foram analisadas 483 amostras de seis diferentes origens. Os espectros foram registrados entre 10.000 e 4000 cm^{-1} , média de 82 varreduras em uma resolução nominal de 4 cm^{-1} , em sementes cortadas ao meio de forma longitudinalmente, no modo de reflectância. Os resultados demonstraram que mais de 95% das amostras de validação foram corretamente classificadas utilizando o PLS-DA. Resultados similares foram obtidos utilizando-se a técnica SIMCA. Os autores concluíram que a associação da espectroscopia NIR e técnicas de classificação pode ser uma valiosa ferramenta para rastrear a origem de pistache, proporcionando uma autenticação confiável de forma rápida e barata.

Santos et al., (2014) avaliando trezentas e cinquenta sementes de duas cultivares de mamona (BRS Nordestina e BRS Paraguaçu) por meio de reflectância difusa e utilizando pré-processamento dos espectros originais e técnicas de reconhecimento de padrões não supervisionado (Análise de Componentes Principais-PCA) e supervisionado (SIMCA) e

aplicação do algoritmo Kennard-Stone para seleção de amostras e a modelagem SPA-LDA concluiu que a PCA permitiu discriminar as cultivares de mamona BRS Nordestina e BRS Paraguaçu na região espectral correspondente à faixa de 2110 a 2155 nm. Assim como o modelo SIMCA forneceu erros de 4 e 1,3% para as classes BRS Nordestina e BRS Paraguaçu nos níveis de significância 1, 5 e 10%, para os conjuntos de validação e teste, classificando 86% das sementes utilizadas no ensaio de campo experimental, o que valida o procedimento utilizado ao passo que O SPA-LDA mostrou-se eficiente, selecionando uma variável espectral classificando corretamente todas as amostras do conjunto teste e 90% das sementes utilizadas no ensaio de campo experimental.

Ênfase vem sendo dada a faixa compreendida entre 780 a 1100nm denominada de região do infravermelho próximo de ondas curtas (SWNIR). De acordo com Barboza (2003) é uma região onde são detectados segundos e terceiros sobretons de estiramento de O-H, N-H e C-H, que apresentam baixa intensidades em relação aos respectivos primeiros sobretons, detectados na região do NIR; ressalta-se ainda que apesar das baixas intensidade, diversos trabalhos indicam uma boa razão sinal/ruído é obtida para esta região.

Algumas das aplicações da espectroscopia SWNIR, são citadas na literatura, dentre elas a determinação de sólidos solúveis em frutas (CARLINI; MASSANTINI; MENCARELLI, 2000; GOLIC; WALSH; LAWSON, 2001) e a determinação de proteínas e lipídios em leite (SASIC; OZAKI, 2001)

Com isso, estudos nesta região espectral, baseados na análise de sementes intactas constitui opção a ser melhor explorada na classificação correta das mesmas, sendo o objetivo deste estudo o desenvolver metodologia para identificação de sementes de mamona de cinco genótipos sendo quatro cultivares de mamona (BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela) e uma linhagem (CNPAM 2009-7)

2. MATERIAL E MÉTODOS

2.1. Aquisição das Amostras

Foram utilizados cinco materiais sendo quatro cultivares de mamona (BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela) e uma linhagem (CNPAM 2009-7), neste trabalho (**Figura 6**).

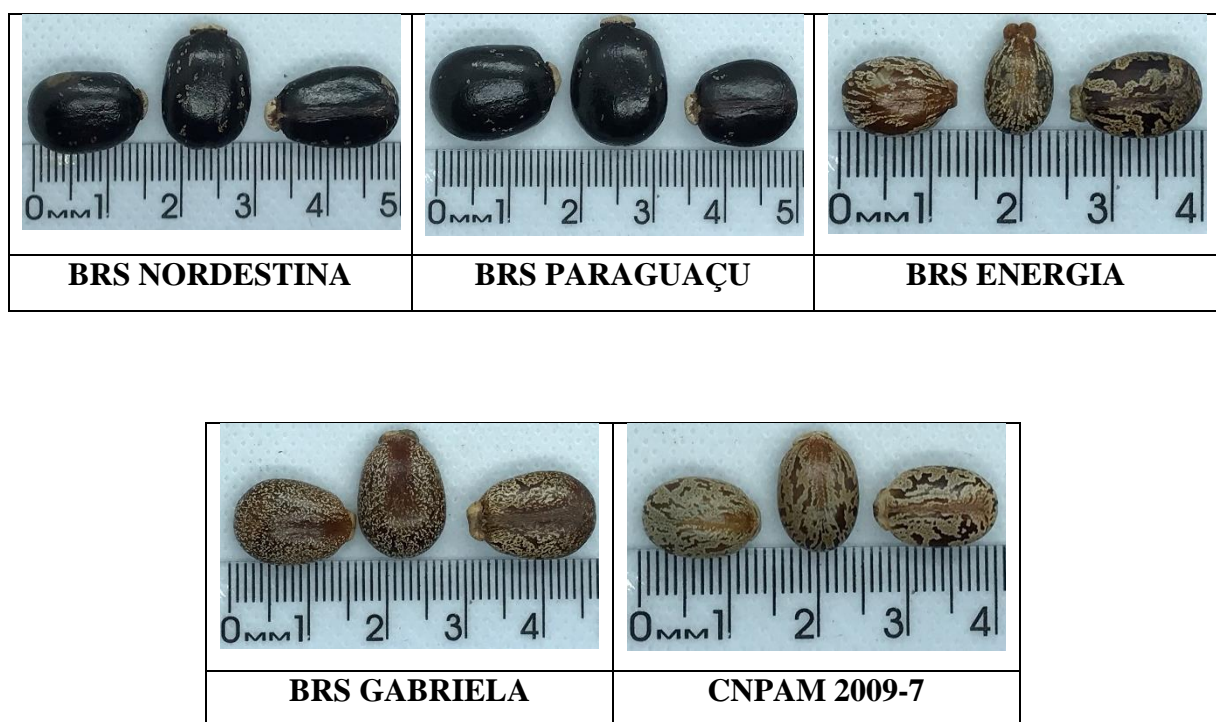


Figura 6- Sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.

Para cada material, utilizou-se cento e vinte e cinco sementes de alta qualidade genética cedidas pela Embrapa Algodão, sendo pesadas e tomadas as suas dimensões (**Tabela 6**), em seguida, foram selecionadas 50 sementes representativas de cada material e acondicionadas a uma temperatura de 21°C e umidade relativa de 70% para realização das avaliações espectrais.

Tabela 6. Caracterização morfológica das sementes de mamona

Cultivar/ Linhagem	Comprimento (mm)	Largura (mm)	Espessura (mm)	Massa (g)	Largura/ Comprimento
BRS Paraguaçu	$17,3 \pm 0,7$	$13,3 \pm 0,5$	$7,3 \pm 0,2$	$0,8 \pm 0,1$	$0,77 \pm 0,01$
BRS Energia	$13,2 \pm 0,6$	$8,6 \pm 0,3$	$6,4 \pm 0,3$	$0,4 \pm 0,0$	$0,66 \pm 0,03$
BRS Gabriela	$15,2 \pm 0,7$	$10,9 \pm 0,4$	$6,3 \pm 0,2$	$0,5 \pm 0,1$	$0,72 \pm 0,02$
BRS Nordestina	$16,0 \pm 0,6$	$11,6 \pm 0,4$	$6,8 \pm 0,2$	$0,6 \pm 0,1$	$0,72 \pm 0,02$
CNPAM 2009-7	$12,9 \pm 0,5$	$8,8 \pm 0,3$	$6,3 \pm 0,2$	$0,3 \pm 0,0$	$0,68 \pm 0,02$

2.2. Instrumentação

As medidas de reflectância difusa foram obtidas em um espectrofotômetro VIS-NIR modelo XDS Rapid Content™ Analyser (Foss Analytical, Hogans, Sweden) conforme ilustrado na **figura 7(A)**. Na **figura 7B** é possível observar no detalhe o ambiente de leitura da amostra.

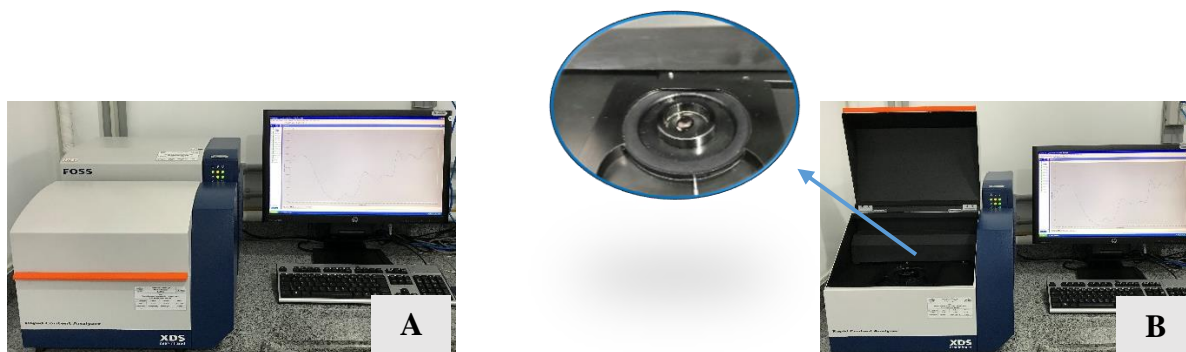


Figura 7- (A) Espectrofotômetro VIS-NIR; (B) Detalhe ambiente de leitura da amostra

Na **figura 8 (A)** observa-se a célula de quartzo circular de 3 cm de diâmetro usada para posicionar a amostra a ser analisada. Para bloquear a radiação espúria do ambiente, foram usadas tampas reflexivas na célula de amostragem, (**Figura 8 B**).

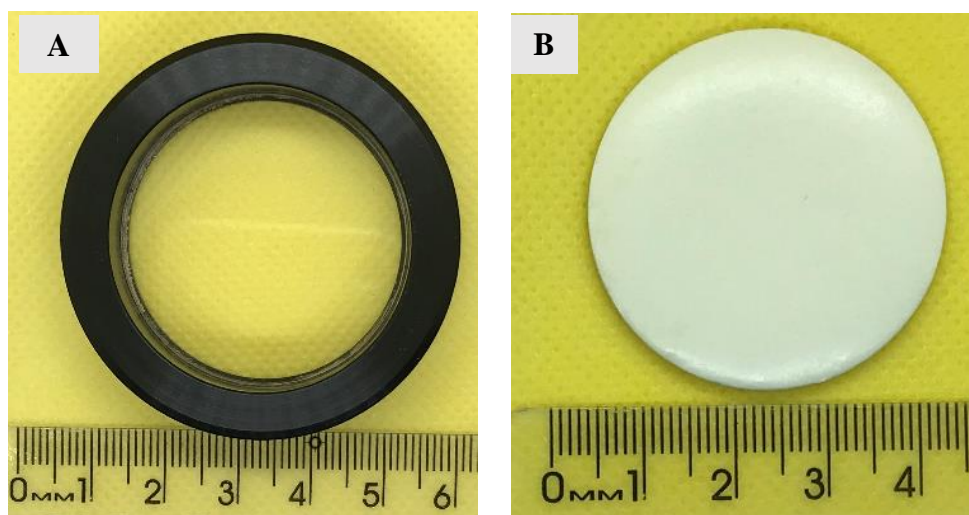


Figura 8- (A) Célula de quartzo; (B) Tampas reflexivas para a célula de quartzo.

Inicialmente com o intuito de registrar o sinal de base utilizou-se um padrão de refletância conforme observado na **Figura 9**.



Figura 9- Padrão de refletância

2.3. Aquisição dos espectros NIR

Os espectros de refletância foram obtidos diretamente sem nenhum tratamento químico das sementes. As amostras sempre foram dispostas na célula, da mesma maneira (com a rafia voltada para baixo) para assegurar a uniformidade das medidas. Cada espectro foi obtido a partir de 32 varreduras na faixa de 400 a 2500nm em intervalos de 0,5nm.

2.4. Programas Computacionais

O pré-processamento dos espectros originais e a aplicação das técnicas de reconhecimento de padrões não supervisionado (Análise de Componentes Principais-PCA) e supervisionado (SIMCA) foram realizados utilizando-se o programa Unscrambler® 9.8. A aplicação do algoritmo Kennard-Stone utilizado para seleção de amostras e a modelagem SPA-LDA foram realizadas em ambiente do software Matlab R2008a.

2.5. Tratamento Quimiométrico dos Dados

2.5.1. Pré-processamento

A região espectral de 750 a 1100nm foi selecionada a priori como a região de trabalho para classificação de sementes de mamoneira que corresponde a região do infravermelho próximo de ondas curtas (SWNIR).

As técnicas de suavização Savitzky-Golay, correção multiplicativa de sinais (MSC), correção de linha de base e derivação, foram avaliadas no pré-processamento dos espectros.

2.5.2. Reconhecimento de Padrões

Realizou-se uma análise exploratória utilizando-se a PCA com o objetivo de observar a formação de agrupamentos.

O algoritmo Kennard-Stone foi aplicado separadamente aos espectros de cada cultivar de mamona com a finalidade de dividir as amostras em conjuntos de treinamento, validação e teste (**Tabela 7**). Esses conjuntos foram utilizados na análise de classificação SIMCA e na modelagem SPA-LDA.

Tabela 7. Número de amostras dos conjuntos de validação cruzada, selecionadas pelo algoritmo KS, para as classes BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e uma linhagem CNPAM 2009-7.

Cultivar/	Conjuntos		Total
	Validação cruzada	Predição	
BRS Paraguaçu	36	14	50
BRS Energia	36	14	50
BRS Gabriela	36	14	50
BRS Nordestina	36	14	50
CNPAM 2009-7	36	14	50

3. RESULTADOS E DISCUSSÃO

3.1. Espectros NIR

Buscando uma melhor interpretação e avaliação dos resultados obtidos e devido as similaridades morfológicas observadas nos materiais analisados, foi feita a separação em dois grupos de trabalho sendo (A) BRS Nordestina e BRS Paraguaçu (cor única) e (B) BRS Energia, CNPAM 2009-7 e BRS Gabriela (Rajadas ou pintadas) (**Figura 10**)



Figura 10- Grupos estabelecidos (A) BRS Nordestina (I) e BRS Paraguaçu (II) e grupo (B) BRS Energia (I), linhagem CNPAM 2009-7(II) e BRS Gabriela III).

Na **figura 11** são observados os espectros originais de 250 amostras das diferentes sementes de mamoneira: BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e a linhagem

CNPAM 2009-7 obtidos na faixa espectral de 400 a 2500nm. Na **figura 12**, foi selecionada apenas a faixa do SWNIR (750-1100) que corresponde a região de trabalho neste estudo.

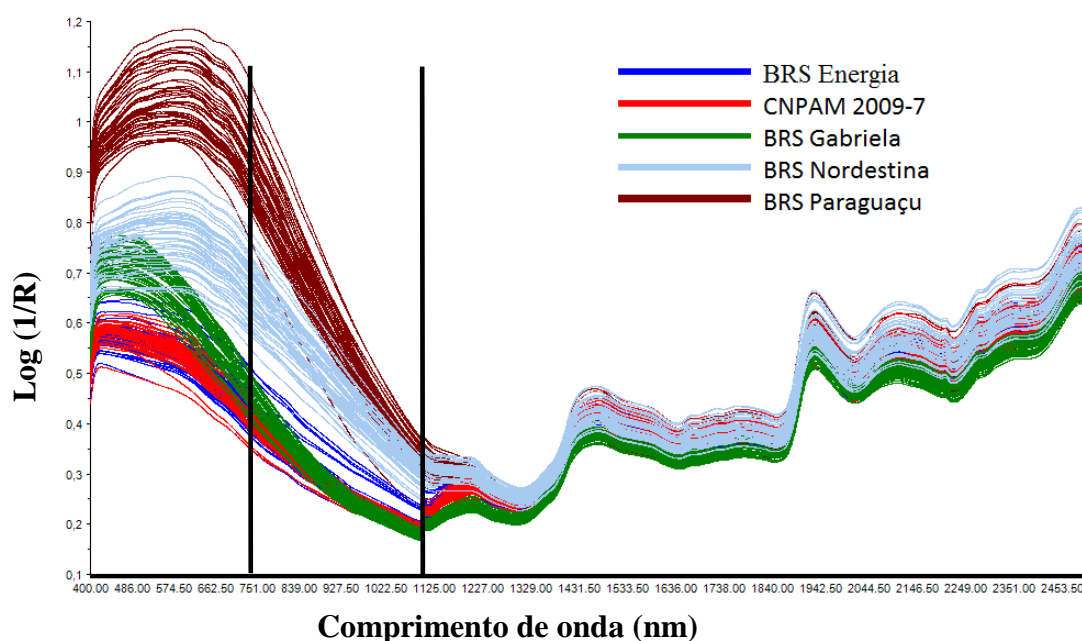


Figura 11- Espectros originais NIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7, na faixa de 400-2500nm.

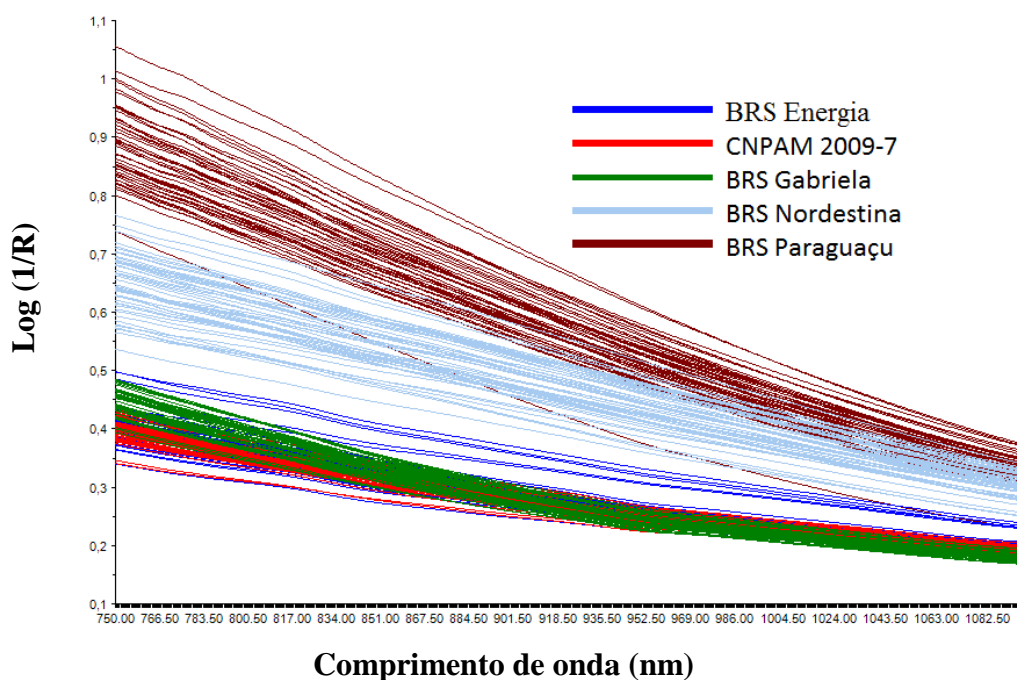


Figura 12- Espectros originais SWNIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7

No gráfico original NIR foi observada alteração na linha de base, sendo necessário a correção utilizando a primeira derivada Savitzky-Golay com janela de 11 pontos e polinômio de segunda ordem (**Figura 13**)

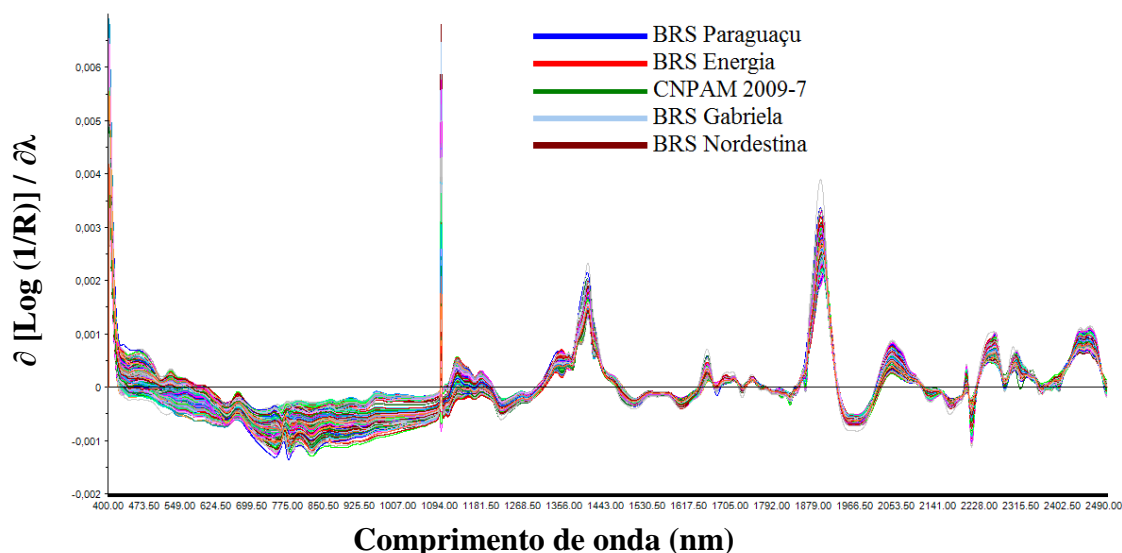


Figura 13- Espectros NIR derivativos das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7, na faixa de 400-2500nm.

Em seguida para melhor visualização dos genótipos em relação ao gráfico foram calculados os espectros médios SWNIR (**Figura 14**).

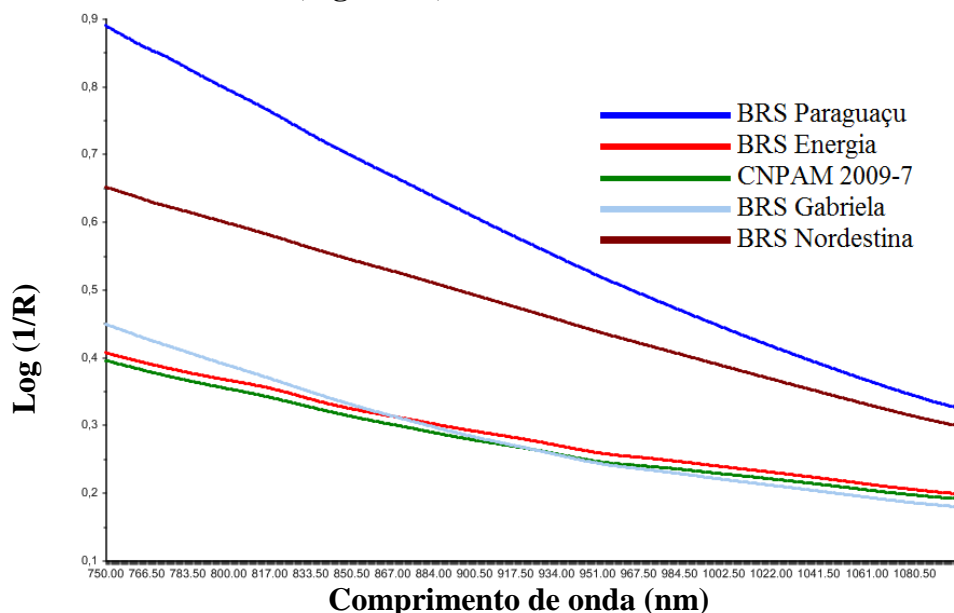


Figura 14- Espectros originais médios SWNIR de reflectância difusa das sementes de mamona, BRS Nordestina e BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.

Mesmo após a seleção dos espectros médios foi identificado um ruído instrumental necessitando que os mesmos sofressem pré-processamento. Os espectros médios derivativos das 250 amostras de sementes de mamoneira são visualizados na **figura 15** na faixa NIR e na **figura 16** na faixa do SWNIR, observando-se a correção do incremento de linha de base com o procedimento empregado.

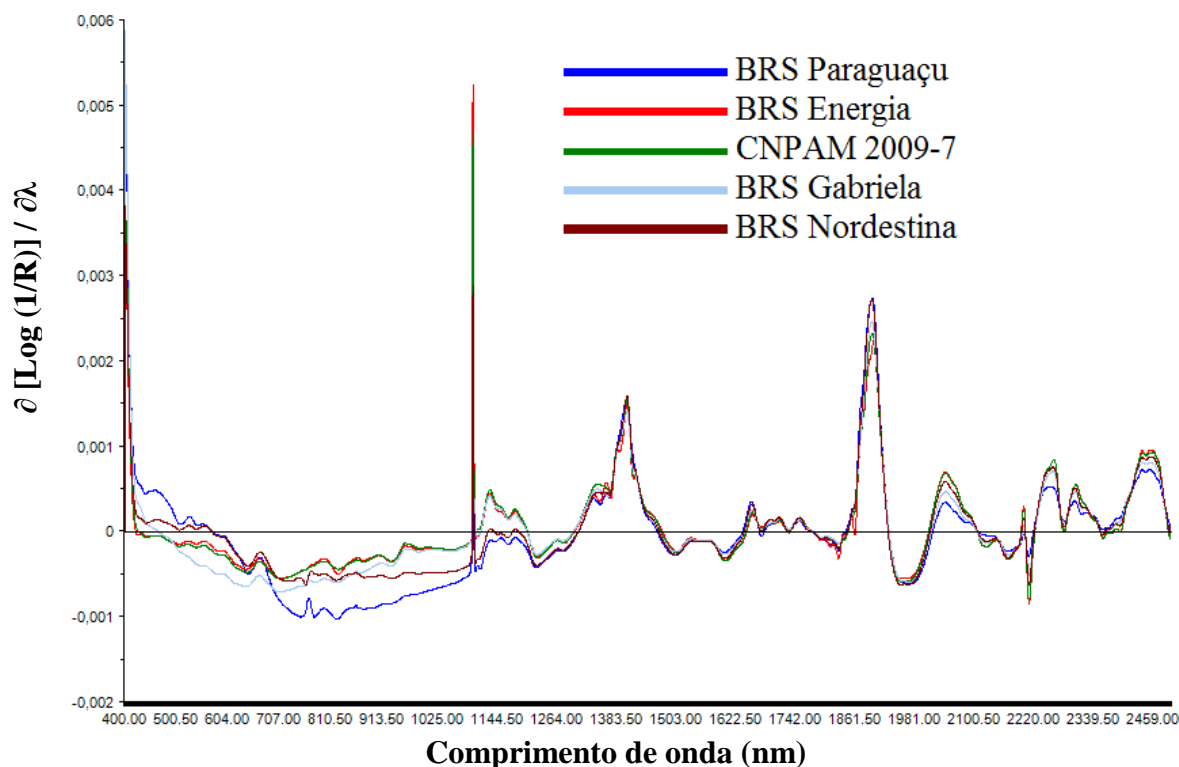


Figura 15- Espectros médios NIR de reflectância difusa pré-processados das sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7

Após o pré-processamento é possível identificar em algumas das amostras a existência de separação, ao passo que, áreas de sobreposição são identificadas.

Essas sobreposições se devem provavelmente ao fato das amostras possuírem bases espectrais muito próximas dificultando a distinção e melhor visualização entre os materiais em estudo.

Ao serem isolados os espectros médios da faixa SWNIR, é possível observar nítida separação nas seguintes faixas: 770-815nm; 820-850nm e 960-1025nm (**Figura 16**).

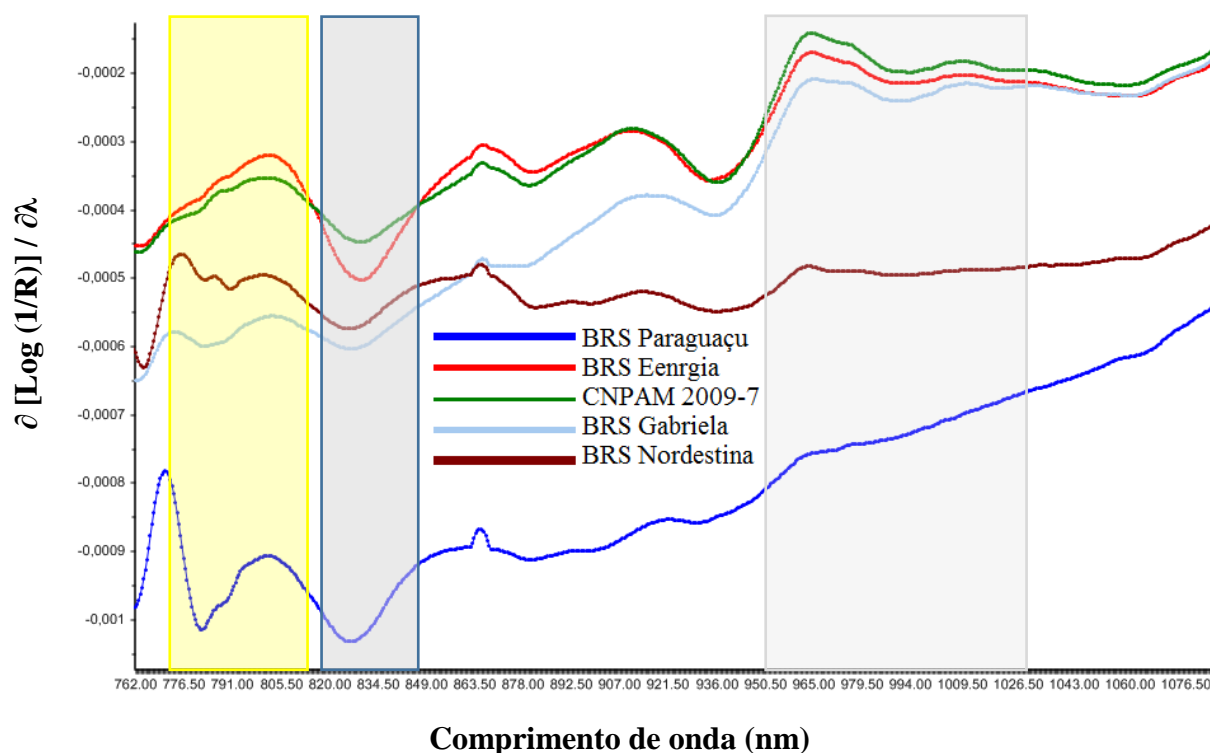


Figura 16- Espectros médios SWNIR de reflectância difusa pré-processados das sementes de mamona, BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.

Os resultados obtidos e a dificuldade nítida em se realizar uma separação visual das sementes de alguns materiais seja por avaliação morfológica, colorimétrica e química, aliada a uma base genética estreita e a complexidade do sinal obtido para as sementes de BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7, é imprescindível o uso de ferramentas quimiométricas para realização do estudo.

3.2. Análise Exploratória dos Dados

Na **figura 17** observa-se o gráfico dos escores referentes aos espectros pré-processados dos genótipos estudados, como resultado da aplicação da PCA, sendo no eixo X a PC_1 e no Y a PC_2 .

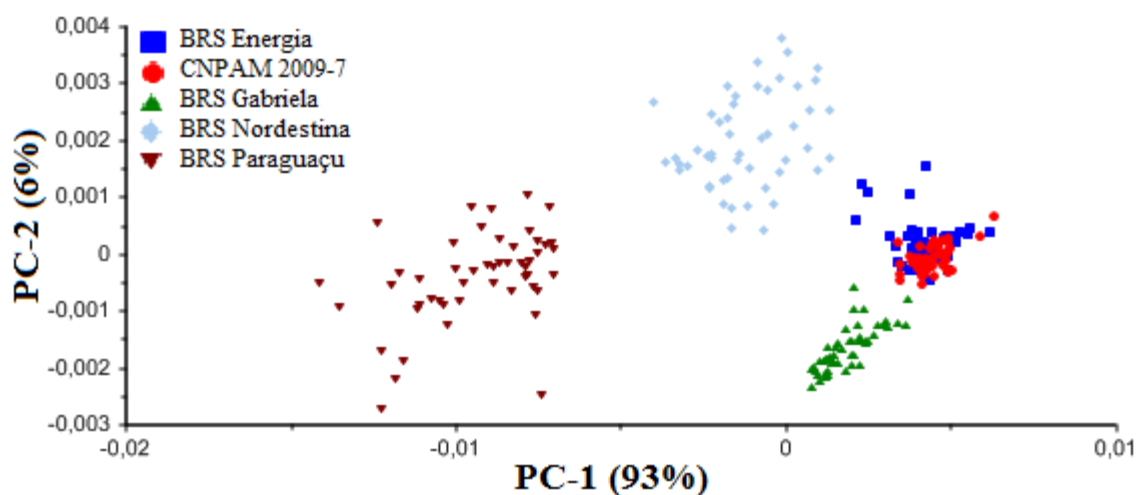


Figura 17- Gráfico dos escores (PC1 vs PC2) para o conjunto das 250 amostras de sementes de mamona BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.

Observa-se por intermédio dos gráficos que a distinção entre os genótipos de mamona ocorre em quase sua totalidade na PC₁, no entanto, se faz necessário identificar quais os principais comprimentos de onda responsáveis por essa distinção, sendo examinado no gráfico de pesos da PC1 (**Figura 18**).

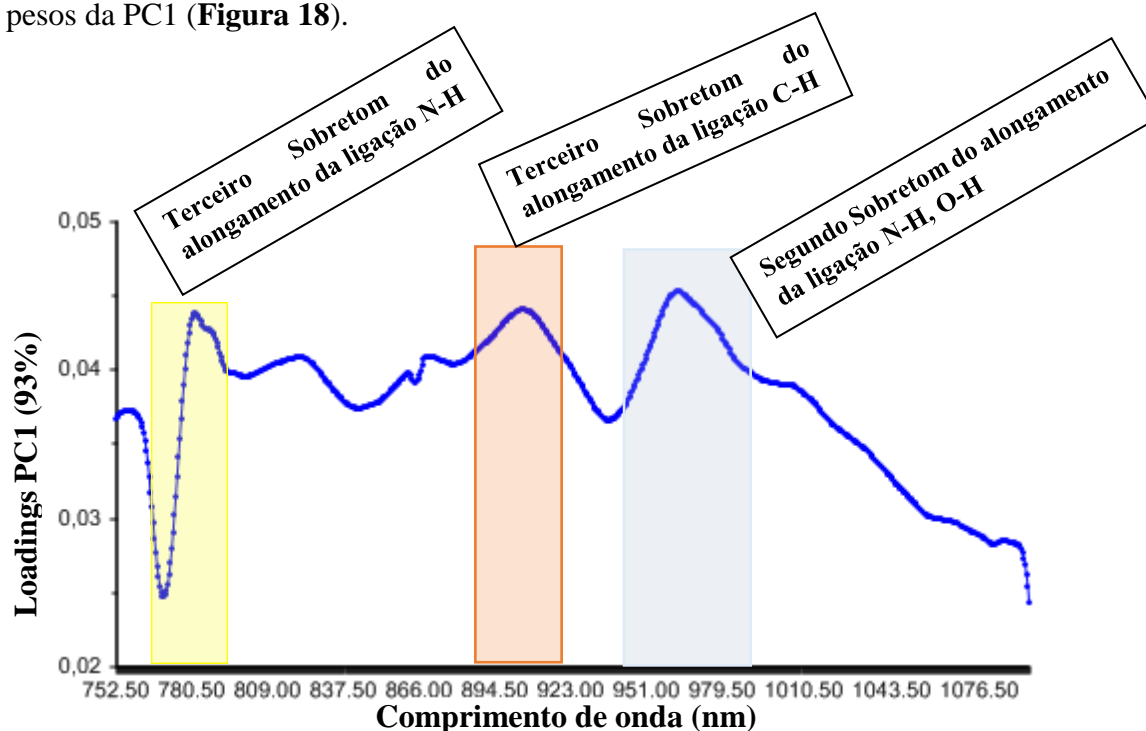


Figura 18- Gráfico de pesos na PC1 para o conjunto das 250 amostras de sementes de mamona BRS Nordestina, BRS Paraguaçu, BRS Energia, BRS Gabriela e CNPAM 2009-7.

É possível observar na faixa compreendida entre 770-800nm e a de 950-980nm como sendo de picos espectrais mais expressivos, desta forma, essa região poderá se destacar na diferenciação entre os genótipos estudados neste trabalho. Tal fato, pode estar relacionado a diferentes teores ligados ao grupo proteico nos genótipos ou outros componentes relacionados ao alongamento das ligações N-H e O-H.

Na **figura 19** observa-se o gráfico dos escores aplicados ao grupo I envolvendo os genótipos BRS Nordestina e BRS Paraguaçu como resultado da aplicação da PCA.

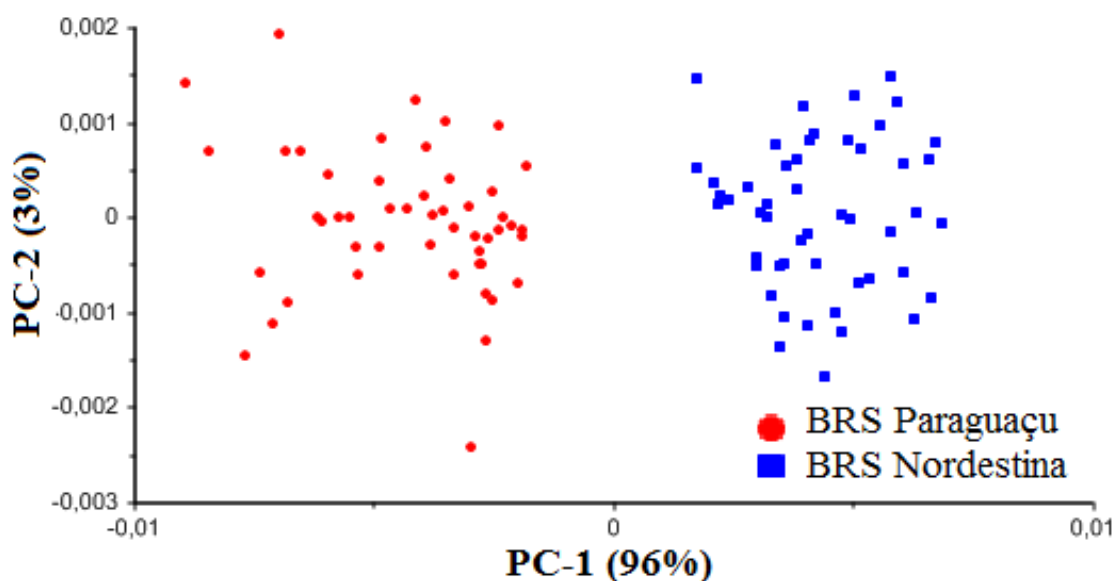


Figura 19- Gráfico dos escores (PC1 vs PC2) para o conjunto das 100 amostras de sementes de mamona das cultivares BRS Nordestina e BRS Paraguaçu.

Mediante os dados visualizados no gráfico de escores é possível a separação entre as cultivares BRS Nordestina e BRS Paraguaçu, não sendo visualizada nenhuma sobreposição o que demonstra o excelente desempenho dos modelos de classificação.

Observa-se, ainda, no gráfico dos escores, que não existem amostras isoladas e, portanto, as 100 amostras foram utilizadas.

A separação entre os genótipos verificada mediante a análise das sementes ocorre praticamente ao longo da PC₁ com uma variância explicada de 96%. Para melhor compreender ao longo da PCA em quais comprimentos de ondas esse efeito é melhor explicado foi avaliado o gráfico de pesos de PC₁ (**Figura 20**).

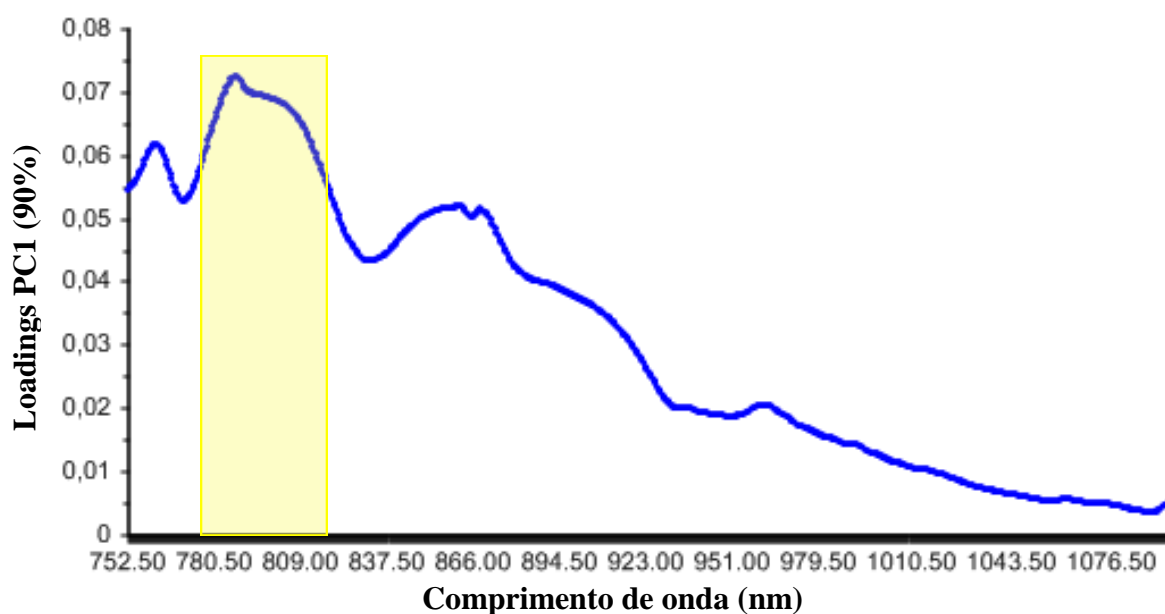


Figura 20- Gráfico de pesos (PC1) para o conjunto das 100 amostras de sementes de mamona BRS Nordestina e BRS Paraguaçu.

É possível observar na faixa compreendida entre 780-820nm como sendo de maiores alterações espectrais, desta forma, essa região que corresponde ao segundo sobretom do alongamento da ligação N-H se destaca na diferenciação entre os genótipos BRS Nordestina e BRS Paraguaçu, estudados neste trabalho, estando possivelmente ligado a grupos proteicos ligados as amostras em estudo.

Na **figura 21** observa-se o gráfico dos escores como resultado da aplicação da PCA (PC1 versus PC2) aos espectros pré-processados das cultivares BRS Energia, BRS Gabriela e da linhagem CNPAM 2009-7.

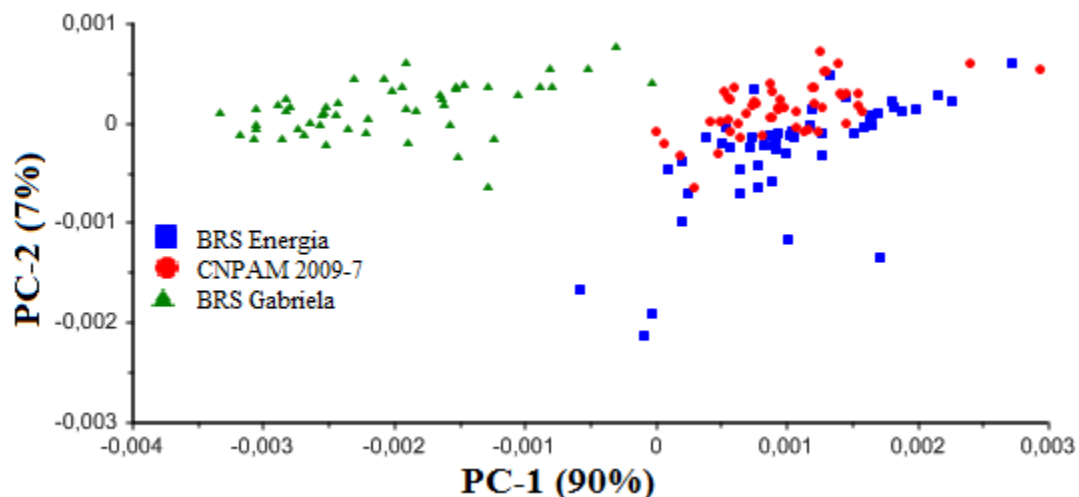


Figura 21- Gráfico dos escores (PC1 vs PC2) para o conjunto das 150 amostras de sementes de mamona BRS Energia, BRS Gabriela e CNPAM 2009-7.

Com base no gráfico de escores há uma tendência de separação entre as amostras dos genótipos BRS Gabriela em relação a BRS Energia, e a linhagem CNPAM 2009-7 em PC1. Contudo, também ocorre uma sobreposição entre os genótipos BRS Energia e a linhagem CNPAM 2009-7, nesse cenário é possível que o grau de parentesco entre os dois genótipos cuja diferença entre eles corresponde a um único gene pode dificultar essa separação. Sem considerar o grau de parentesco entre os genótipos é possível que haja um comprometimento em relação ao desempenho do modelo de classificação, necessitando de uma melhor avaliação.

Observa-se, ainda, no gráfico dos escores, que não existem amostras isoladas e, portanto, as 150 amostras foram utilizadas.

É possível observar também que a diferenciação dos genótipos apesar da sobreposição verificada ocorre em sua grande maioria ao longo da PC₁. Se faz necessário a identificação dos comprimentos de onda ao longo da referida PC que contribuem com maior ênfase para esta diferenciação (**Figura 22**).

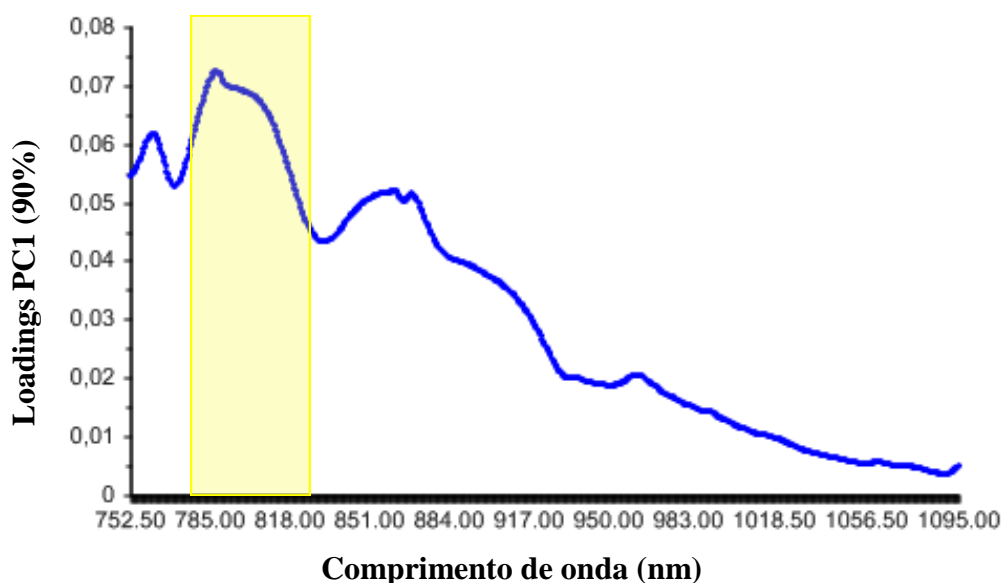


Figura 22- Gráfico de pesos (PC1) para o conjunto das 150 amostras de sementes de mamona BRS Energia, BRS Gabriela e CNPAM 2009-7.

A faixa compreendida entre 780-820nm corresponde a região do espectro que registra picos espectrais de maiores amplitudes, desta forma, essa região poderá se destacar na diferenciação entre os genótipos do grupo II, estudados neste trabalho. Essa diferenciação corresponde a faixa ligada ao terceiro sobretom de alongamento da ligação N-H e poderá estar correlacionada com diferentes teores proteicos associados aos diferentes genótipos estudados

3.3. Reconhecimento de Padrões Supervisionados

3.3.1. Construção e Validação dos Modelos SIMCA

Após a construção das PCAs baseadas em conjuntos completos das amostras e para os dois grupos em estudo, ficou evidenciado que os grupos formados, possuem homogeneidade de comportamento, demonstrando a eficiência da seleção das amostras de validação realizadas com o algoritmo KS.

No total, cinco PCs foram selecionadas para todas as classes estas explicam mais de 99% de todos os modelos. Vale salientar que duas PCs explicam 99% do modelo envolvendo todas as classes, 99% para a classe envolvendo os genótipos BRS Nordestina e BRS Paraguaçu e 97% para a classe envolvendo os genótipos BRS Energia, CNPAM 2009-7 e BRS Gabriela. O número ótimo de PCs na PCA foi usado na classificação SIMCA.

Na **tabela 8** são ilustrados os erros de classificação do conjunto de previsão com o objetivo de avaliar o desempenho dos modelos construídos.

Tabela 8. Resumo de erros de classificação obtido pelos modelos SIMCA utilizando-se o conjunto de amostras de previsão das sementes de mamona nos níveis de significância do Teste – F (1, 5, 10 e 25%).

Modelos SIMCA	% Teste F	Amostras de Previsão				
		P	N	E	C	G
P	1	3	0	-	-	-
	5	3	0	-	-	-
	10	3	0	-	-	-
	25	6	0	-	-	-
N	1	2	0			-
	5	2	2	-	-	-
	10	1	4	-	-	-
	25	0	7	-	-	-
E	1	-	-	0	11	0
	5	-	-	1	6	0
	10	-	-	1	1	0
	25	-	-	5	0	0
C	1	-	-	7	0	0
	5	-	-	1	0	0
	10	-	-	0	0	0
	25	-	-	0	3	0
G	1	-	-	0	1	0
	5	-	-	0	0	0
	10	-	-	0	0	0
	25	-	-	0	0	0

- = não testado em relação ao outro agrupamento

Os resultados de erros de classificação encontrados reforçam a possibilidade de implantação dos modelos, exceto para o modelo com 25% como nível de significância. Neste observa-se 50% erros do **Tipo I** para cultivar BRS Nordestina, 42,8% para cultivar BRS Paraguaçu, 35,7% para a BRS energia, 21,4% para a CNPAM 2009-7 e 0% para a cultivar BRS Gabriela, o que podemos correlacionar a diferenças visuais mais fortemente perceptíveis na semente desta, em relação as dos demais genótipos.

Os erros do **Tipo II** são menos frequentes e destacando-se no cultivar BRS Energia com 43%, este provavelmente é explicado em função da base genética estreita entre a BRS energia e a CNPAM 2009-7, que de acordo com o melhoramento diferem em apenas um gene.

Neste particular, vale considerar a complexidade da matriz e o número de amostras analisadas. Portanto, os modelos foram considerados validados e o nível de significância escolhido foi o de 5%, por apresentar os melhores índices de respostas aos conjuntos analisados e por ser o mais utilizado na literatura (**Tabela 09**).

Tabela 9. Resumo de erros de classificação obtido pelos modelos SIMCA utilizando-se o conjunto de amostras de previsão das sementes de mamona ao nível de significância do Teste – F de 5%.

Modelos SIMCA	Amostras de Previsão				
	P	N	E	C	G
P	3	0	-	-	-
N	2	2	-	-	-
E	-	-	1	6	0
C	-	-	1	0	0
G	-	-	0	0	0

- = não testado em relação ao outro agrupamento

3.3.2. Construção e Validação do Modelo SPA-LDA

O número ideal de variáveis para o SPA-LDA foi determinado a partir do mínimo da função de custo G , exibido na **figura 23**.

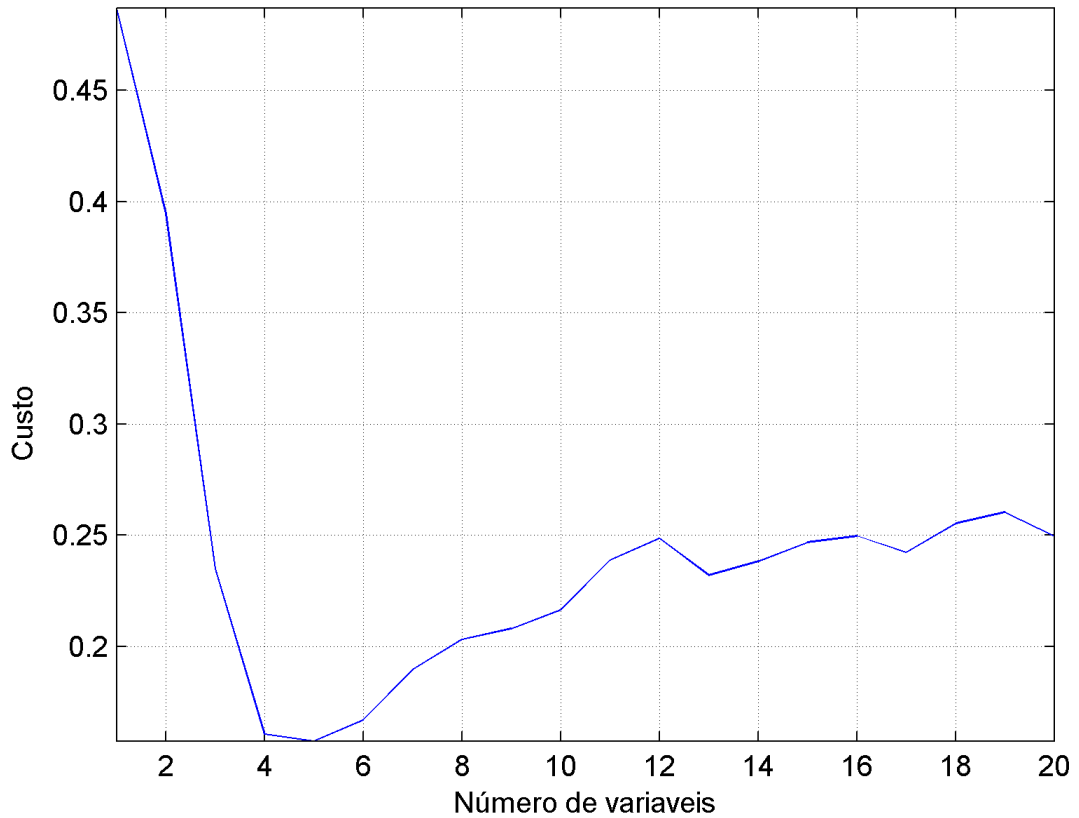


Figura 23- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA

Mediante a aplicação do algoritmo APS (**Tabela 10**), foram selecionadas 5 variáveis como sendo o número ideal não apresentando erros de teste e nem erros de predição, (**Figura 23**) sendo o menor custo obtido no conjunto de teste de 0.15742,

Tabela 10. Resultados do algoritmo SPA-LDA para o conjunto de previsão separado pelo algoritmo KS.

Nome da Amostra	Referência	Classificação pelo SPA-LDA
3 C	1	1
12 C	1	1
14 C	1	1
16 C	1	1
38 C	1	1
44 C	1	1
48 C	1	1
50 C	1	1
58 C	1	1
68 C	1	1
3 E	2	2
32 E	2	2
53 E	2	2
67 E	2	2
68 E	2	2
89 E	2	2
97 E	2	2
102 E	2	2
103 E	2	2
114 E	2	2
47 G	3	3
48 G	3	3
50 G	3	3
57 G	3	3
59 G	3	3
65 G	3	3
78 G	3	3
108 G	3	3
117 G	3	3
124 G	3	3
4 N	4	4
36 N	4	4
40 N	4	4
51 N	4	4
54 N	4	4
64 N	4	4
69 N	4	4
74 N	4	4
98 N	4	4
113 N	4	4
7 P	5	5
38 P	5	5
59 P	5	5
70 P	5	5
74 P	5	5
76 P	5	5
101 P	5	5
109 P	5	5
112 P	5	5
115 P	5	5

Na **figura 24** observa-se o espectro médio das amostras de treinamento ao qual foi indexada as variáveis selecionadas pelo SPA-LDA.

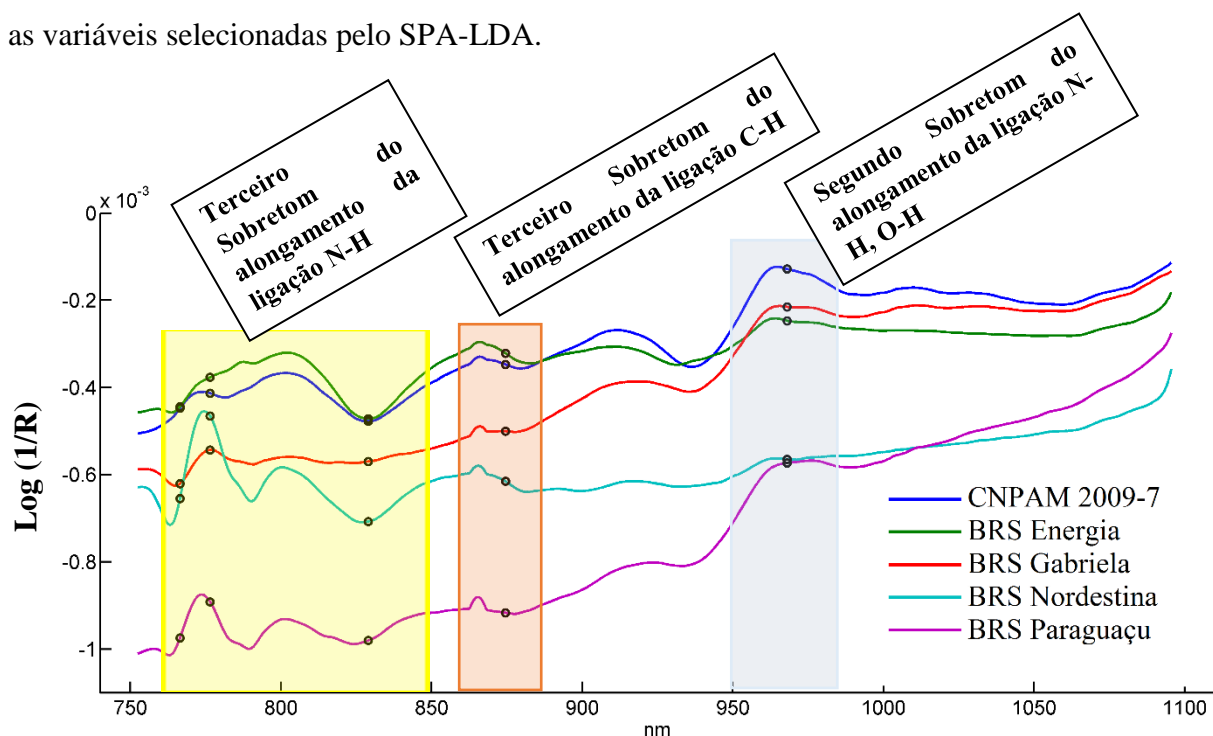


Figura 24- Espectro médio das amostras de treinamento com as variáveis selecionadas pelo SPA-LDA.

Conforme observado na **figura 24**, podemos relacionar após a indexação das variáveis selecionadas aos espectros médios que a separação entre os genótipos estudados ocorre envolvendo compostos químicos ligados ao teor de óleos e proteínas. Vale salientar ainda que em relação ao sobretom relativo ao alongamento da ligação O-H, este pode estar relacionado a diferentes teores de ácido ricinoleico, que pode estar propiciando a diferenciação nos genótipos estudados

3.4. Aplicação dos Modelos ao Conjunto de Teste

Na **tabela 11** são descritos os erros de classificação obtidos pelos modelos SIMCA e SPA-LDA em um conjunto externo como forma de verificar o desempenho de ambos. Os parâmetros utilizados nos modelos SIMCA foram: quatro PCs, para cada classe e 5% de nível de significância estatística.

Tabela 11. Resumo da aplicação dos modelos SIMCA e SPA-LDA no conjunto de teste ao nível de significância de 5%, 5 variáveis latentes para SIMCA e 10 variáveis selecionadas pelo APS-LDA.

CLASSES	MODELOS									
	SIMCA (5)					SPA-LDA (10)				
	P	N	E	C	G	P	N	E	C	G
P	3	2				0	0			
N	0	2				0	0			
E			1	1	0			0	0	0
C			6	0	0			0	0	0
G			0	0	0			0	0	0

A capacidade de discriminação de ambos os modelos foi comprovada no conjunto de previsão. Contudo, o modelo SPA-LDA para esse tipo de matriz, mostrou-se eficaz classificando corretamente todas as amostras do conjunto de teste.

O uso de ambos os modelos no melhoramento de plantas constitui uma ferramenta de extrema importância, haja vista, a indisponibilidade de sementes em quantidades maiores, as técnicas de classificação não necessitem da destruição das sementes e serem efetuadas as classificações de forma rápida.

Mesmo havendo respostas satisfatória para o uso do método SPA-LDA, recomenda-se o s dois modelos, ou seja, o modelo SIMCA associado ao SPA-LDA, para dar maior garantia que não existe classificação equivocada das sementes minimizando ou mesmo eliminando os erros de classificação

4. CONCLUSÕES

- A PCA permitiu discriminar as cultivares de mamona BRS Nordestina, BRS Paraguaçu, na região espectral correspondente à faixa de 750 a 1100nm;
- O modelo SIMCA forneceu erros do tipo I de 21,4% para BRS Nordestina, 14,3% para BRS Paraguaçu, 7,1% para BRS Energia, não sendo registrados erros para CNPAM 2009-7 e BRS Gabriela no nível de significância de 5%, para o conjunto de previsão;
- O SPA-LDA mostrou-se eficiente, selecionando cinco variáveis espectrais classificando corretamente todas as amostras do conjunto de previsão.

- Os modelos utilizados são promissores na classificação de sementes de mamona na faixa espectral do SWNIR, necessitando de maior aprofundamento na determinação de qual ou quais comprimentos de onda são responsáveis pelas melhores respostas;
- Deve se buscar a predição não destrutiva das amostras de componentes químicos presentes na amostra e detectáveis na região SWNIR para auxiliar na tomada de decisão em programas de controle de qualidade em laboratórios de análise de sementes.

5. REFERENCIAS BIBLIOGRAFICAS

ASSOCIATION OF OFFICIAL SEED ANALYSIS. **Seed vigor testing handbook**. East Lansing, 1983. 88p.

BALABIN, R. M.; SAFIEVA, R. Z. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data. **Analytica Chimica Acta**, v. 689, p. 190-197, 2011.

BALABIN, R. M.; SAFIEVA, R. Z.; LOMAKINAC, E. I. Gasoline classification using near infrared (NIR) spectroscopy data: Comparison of multivariate techniques. **Analytica Chimica Acta**, v. 671, p. 27-35, 2010.

BASTOS, L. F.; ALESSIO, K. O.; MENEZES, C. M.; COSTA, A. B. da.; MARDER, L. Aplicação da espectroscopia no infravermelho próximo e da análise exploratória de dados para identificação de madeiras na indústria de móveis, **TECNO-LÓGICA**, v. 16, n. 1, p. 25-29, jan./jun. 2012.

BELTRÃO, N. E. M.; AZEVEDO, D. M. P. Fitologia. In: AZEVEDO, D. M. P. de; BELTRÃO, N. E. de M. (Eds). **O agronegócio da mamona no Brasil**. 2. ed. Brasília: Embrapa Comunicação para Transferência de Tecnologia, 2007. cap. 5, p.119-137.

BETEMPS, D. L.; FACHINELLO, J. C.; GALARÇA, S. P.; MACHADO, N. P.; REMORINI, D.; MASSAI, R.; AGATI, G., Espectroscopia do visível e infra-vermelho próximo para estimar

sólidos solúveis e firmeza de polpa em função da época de colheita em pêssegos. **Semina: Ciências Agrárias**, v. 35, n. 3, p. 1257-1266, maio/jun. 2014.

BRASIL. Ministério da Agricultura Pecuária e Abastecimento. **Regras para análise de sementes**. Brasília: Secretaria Nacional de Defesa Agropecuária, 2009. 399 p.

BRETTING, P. K.; WIDRLECHENER, M. P. Genetic markers and horticultural germplasm management. **HortScience**, v. 30, n. 7, p. 1349-1355, 1995.

CARLINI, P.; MASSANTINI, R.; MENCARELLI, F. Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. **Journal of Agricultural and Food Chemistry**, v. 48, p. 5236-5242, 2000.

CHEN, L.; WANG, J.; YE, Z.; ZHAO, J.; XUE, X.; HEYDEN Y.; SUN, Q. Classification of Chinese honeys according to their floral origin by Near Infrared Spectroscopy. **Food Chemistry**, v. 135, p. 338-342, 2012.

DINIZ, M. R.; RIFFEL, R. A., Uma biblioteca de espectros estelares no infravermelho próximo para medidas de cinemática estelar em galáxias. **Ciência e natureza - UFSM**, v. 33 (2), p. 43-56, 2011.

EGIDIO, V.; OLIVERI, P.; WOODCOCK, T.; DOWNEY, G. Confirmation of brand identity in foods by near infrared transreflectance spectroscopy using classification and class-modelling chemometric techniques — The example of a Belgian beer. **Food Research International**, v. 44, p. 544-549, 2011.

FERRARESI, T. M.; SILVA, W. T. L. da.; MARTIN-NETO, L.; SILVEIRA, P. M. da.; MADARI, B. E., Espectroscopia de infravermelho na determinação da textura do solo. **Revista Brasileira de Ciência do Solo**, v. 36, p.1769-1777, 2012.

FERREIRA, M. A. J. da F. **Utilização das técnicas de marcadores moleculares na genética de populações, na genética quantitativa e no melhoramento de plantas**. Boa Vista: Embrapa Roraima, 2003. 63p. Documento, 1.

FOLHA, T. O. **Uso da espectroscopia no infravermelho próximo (NIR) para a avaliação de parâmetros de qualidade de farinha de mandioca**. 101 f.: fig., tab. Dissertação (Mestrado) - Universidade Federal de Pernambuco. CCEN. Química Fundamental, 2014.

GALTIER, O.; ABBAS, O.; LE DREAU, Y.; REBUFA, C.; KISTER, J.; ARTAUD, J.; DUPUV, N.. Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. **Vibrational Spectroscopy**, v. 55, p. 132-140, 2011.

GHASEMI-VARNAMKHAHI, M.; MOHTASEBI, A. S.; RODRIGUEZ-MENDEZ, M. L.; GOMES, A. A.; ARAUJO, M. C. U.; GALVÃO, R. K. H. Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection. **Talanta**, v. 89, p. 286-291, 2012.

GOLIC, M.; WALSH, K.; LAWSON, P. “**10th International Conference on Near Infrared Spectroscopy**”, Korea, 2001.

HORN, M. B., **Micotoxinas em Silagens de Milho do Sul do Brasil e Metodologia Analítica para Aflatoxinas por Espectroscopia de Infravermelho Proximo em Milho**. Dissertação (mestrado) – Universidade Federal de Santa Catarina, Centro de Ciências Agrárias – Programa de Pós-Graduação em Ciência dos Alimentos. 2013. 122 p.

INSAUSTI, M.; GOMES, A. A.; CRUZ, F. V.; PISTONESI, M. F.; ARAUJO, M. C.; GALVÃO, R. K.; PEREIRA, C. F.; BAND, B. S. Screening analysis of biodiesel feedstock using UV-vis, NIR and synchronous fluorescence spectrometries and the successive projections algorithm. **Talanta**, v. 97, p. 579-583, 2012.

INTERNATIONAL SEED TESTING ASSOCIATION. **Handbook of vigour test methods**. Zurich, 1995. 117 p.

JIANG, H. Y.; ZHU, Y. J.; WEI, L. M.; DAI, J. R.; SONG, T. M.; YAN, Y. L.; CHEN, S. J. Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance

spectroscopy (NIRS) in maize (*Zea mays* L.). **Plant Breeding**, Berlin, v. 126, n. 5, p. 492-497, Oct. 2007. DOI. 10.1111/j.1439-0523.2007.01338.x.

LEE, J. H.; CHOUNG, M. G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. **Food Chemistry**, v.126, p. 368-373, 2011.

LIMA, K.; M. G.; RAIMUNDO JUNIOR, I. M.; SILVA A. M. S.; PIMENTEL, M. F., Sensores ópticos com detecção no infravermelho próximo e médio. **Química Nova**, v. 32, p. 1635-1648, 2009

MARQUETTI, I., **Classificação de genótipos de café arábica usando espectroscopia de infravermelho próximo**. Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Tecnologia de Alimentos. Medianeira. 80 f. : il. 2014.

MOREIRA, E. D. T.; PONTES, M. J. C.; GALVÃO, R. K. H.; ARAUJO, M. C. U. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. **Talanta**, v. 79, p. 1260-1264, 2009.

OSBORNE, B. G. Applications of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programmes. **Journal of Near Infrared Spectroscopy**, Sussex, v. 14, n. 2, p. 93-101, 2006. DOI: 10.1255/jnirs.595.

MOSHKIN, V. A. **Castor**. New Delhi: Amerind, 1986. 315 p.

PADILHA, L.; GUIMARÃES, C.T.; VIEIRA, M.G.G.C.; SOUZA, I.R.P.; PARENTONI, S.N., PACHECO, C.A.P., SANTOS, M.X., GAMA, E.E.G.; PAIVA, E. Microsatélites fluorescentes na diferenciação de linhagens de milho. In: CONGRESSO NACIONAL DE MILHO E SORGO, 24., 2002, Sete Lagoas. **Anais...** Sete Lagoas: ABMS/EMBRAPA Milho e Sorgo, 2002. p. 1-5

PRIOLLI, R.H.G.; MENDES-JUNIOR, C.T.; ARANTES, N.E.; CONTEL, E.P.B. Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, v. 25, p. 185-193, 2002.

RIBEIRO. J. S.; SALVA. T. J. G., **Proposta de metodologia para o estabelecimento de modelo de calibração multivariado para a previsão do teor de trigonelina em café cru baseado em espectroscopia no infravermelho próximo**. VII Simpósio de Pesquisa dos Cafés do Brasil 22 a 25 de Agosto de 2011, Araxá – MG.

RIOVANTO, R.; CYNKAR, W.U.; BERZAGHI, P.; COZZOLINO, D. Discrimination between Shiraz Wines from Different Australian Regions: The Role of Spectroscopy and Chemometrics. **Journal of Agricultural and Food Chemistry**, v. 59, p. 10356-10360, 2011.

ROCHA, O. R. **Aplicação da espectroscopia de infravermelho próximo para a determinação do caseino macropéptídeo em leite UAT**. Dissertação (Mestrado) – Universidade Federal de Goiás, Escola de Veterinária, 63 f.: il., figs., tabs. 2010.

SANTOS, G. A.; SANTOS, A. P. dos; KORNDÖRFER, G. H., Sistema por infravermelho próximo (NIR) para análises de nitrogênio foliar, **Bioscience Journal**, v. 28, Supplement 1, p. 83-90, Mar. 2012.

SANTOS, M. B. H., GOMES, A. A., VILAR, W. T. S., ALMEIDA, P. B. A., MILANI, M., NÓBREGA, M. B. M., MEDEIROS, E. P., GALVÃO, R. K. H., ARAÚJO, M. C. U. Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. **Journal Brazilian Chemistry Society**, v.25, p. 969-974, 2014.

SASIC, S.; OZAKI, Y. **Analytical Chemistry**, 73 (2001) 5236

SAVY FILHO, A. **Mamona Tecnologia Agrícola**. Campinas: EMOPI, 2005. 105 p.

SEGNINI. A.; SILVA, W.T.L. da.; AISSA, N.L.; OTAVIANI-JUNIOR, P.L., OLIVEIRA, P.P.A.; MARTIN NETO, L., MILORI, D.M.B.P. Determinação de propriedades do solo utilizando espectroscopia de infravermelho próximo, **SIAGRO-2014**. p. 241-244

SILVA, D. A. da.; ALMEIDA, V. C.; VIANA, L. C.; KLOCK, U.; MUÑIZ , G. I. B. de, Avaliação das Propriedades Energéticas de Resíduos de Madeiras Tropicais com Uso da Espectroscopia NIR. **Floresta e Ambiente**; v.21(4), p.561-568, 2014.

SINELLI, N.; CASALE, M.; EGIDIO, V.; OLIVERI, P.; BASSI, D.; TURA, D.; CASIRAGHI, E. Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. **Food Research International**, v.43, p. 2126-2131, 2010.

TALLADA, J. G.; PALACIOS-ROJAS, G. N.; ARMSTRONG, P. R. Prediction of maize seed attributes using a rapid single kernel near infrared instrument. **Journal of Cereal Science**, London, v. 50, n. 3, p. 381-387, Nov. 2009. DOI: 10.1016/j.jcs.2009.08.003.

VECCHIA, P. T. D.; SILVA, C.A.R.; SOBRINHO TERCENCIANO P. Use of molecular marker techniques in seed testing by brazilian seed companies. **Scientia Agricola**, v. 55, p. 79-81, 1998.

VÉRAS, G.; GOMES, A. A.; SILVA, A. C.; BRITO, A. L. B.; ALMEIDA, P. B. A.; MEDEIROS, E .P. Classification of biodiesel using NIR spectrometry and multivariate techniques. **Talanta**, v.83, p. 565-568, 2010.

VIDAL, M. S.; MILANI, M.; MENEZES, C. H. S. G.; BEZERRA, C. S. **Seleção de marcadores do tipo Rapd para caracterização genética *Ricinus communis* L.** Campina Grande: Embrapa Algodão, 2005. 5p. Documento, 90.

VITALE, R.; BEVILACQUA, M.; BUCCI, R.; MAGRI, A. D.; MAGRI, A. L.; MARINI, F. A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 121, p. 90-99, 2013.

CAPITULO II
MARCADORES
ESPECTRAIS VIS-NIR EM
PROGENIES DE MAMONA

FERREIRA, D.F. **Marcadores espectrais VIS-NIR para seleção em progênies de mamona (*Ricinus communis* L.)**. 2015. 134f. Tese (Doutorado em Agronomia), Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba.

RESUMO

No melhoramento da mamona (*Ricinus communis* L.) a busca em se conhecer a expressão dos caracteres agronômicos, bem como, essas características, as quais, são transferidas ao longo das gerações, desperta o interesse de melhoristas dos principais países produtores. Em programas tradicionais de melhoramento são necessários entre 10 a 12 anos para se ter respostas consolidadas. Dessa forma, o objetivo deste estudo é estabelecer marcadores espectrais de identificação de genótipos de mamona baseado em progenitores e gerações recíprocas. Foram utilizados dois progenitores e quatro gerações recíprocas (F1, F2, F3 e F4) obtidas por intermédio de autofecundação em dois anos de trabalho. Para cada material foram empregadas 20 sementes de alta qualidade genética cedidas pela Embrapa Algodão, sendo estas acondicionadas a uma temperatura de 21°C e umidade relativa de 70% para realização das avaliações. As medidas de reflectância difusa foram obtidas em um espectrômetro VIS-NIR modelo XDS Rapid Content™ Analyser (Foss Analytical, Hogans, Sweden). Os espectros de reflectância foram obtidos diretamente sem nenhum tratamento químico das sementes e registrados na faixa de 400 a 750nm (região do visível VIS-NIR). Para os modelos de classificação foram utilizadas 14 sementes de cada material, e para o conjunto de teste 6 sementes de cada. As leituras foram efetuadas com as sementes sempre na mesma posição. Os espectros foram pré-processados utilizando o algoritmo Savitzky-Golay com janela de 11 pontos e primeira derivada para correção da linha de base. Com base na PCA (Principal Component Analysis) foi observado que a faixa em estudo permitiu a distinção entre os progenitores e também em algumas gerações recíprocas. O modelo SIMCA (Soft Independent Modelling of Class Analogy) demonstrou resultados promissores para classificação nos níveis de significância de 1, 5 e 10%. O SPA-LDA (Sucessive Projections Algorithm-Linear Discriminant Analysis) foi eficiente classificando corretamente todas as amostras em suas classes. Os métodos em estudo foram eficientes permitindo a diferenciação fenotípica para progenitores e gerações recíprocas de mamoneira de forma rápida, pratica e não destrutiva.

Palavras-chave: Produção de sementes, *Ricinus communis* L., Espectroscopia VIS-NIR, Gerações recíprocas, SIMCA, SPA-LDA

FERREIRA, D.F. **VIS-NIR spectral markers for selection in progenies castor bean (*Ricinus communis* L.)**. 2015. 134f. Thesis (Doctorate in Agronomy Universidade Federal da Paraíba, Centro de Ciências Agrárias, Areia, Paraíba, Areia, Paraíba.

ABSTRACT

In breeding of the Castor bean (*Ricinus comunnis* L.) the search for know the expression of the agronomic characters, as well these, characteristics are transferred throughout generations, has aroused interest of breeders of the main producing countries. In traditional breeding programs are needed between 10 to 12 years to have consolidated responses. Thus, the objective of this study is to establish spectral markers for identification of genotypes based on castor progenitors and reciprocal generations. We used two progenitors and four reciprocal generations (F1, F2, F3 and F4) obtained through self-fertilisation in two years of work. For each material were employed 20 high quality genetic seed provided by Embrapa cotton, these being stock at a temperature of 21° C and 70% relative humidity for carrying out the evaluations. The diffuse reflectance measurements were obtained in a spectrometer VIS-NIR ContentTM model XDS Rapid Analyzer (Foss Analytical, Hogans, Sweden). The reflectance spectra were obtained directly without any chemical treatment of the seeds. The reflectance spectra were recorded in the range of 400 to 750nm (visible region VIS-NIR). Classification models were used 14 seeds of each material, and for the test set 6 seeds of each. The readings were made with the seeds always in the same position. The spectra were preprocessed using the algorithm Savitzky-Golay with window of 11 points and first derivative for baseline correction. Based on the PCA (Principal Component Analysis) it was observed that the range in study allowed the distinction between progenitors and also between in some reciprocal generations. The SIMCA model (Soft Independent Modelling of Class Analogy) demonstrated promising results for classification in the significance levels 1, 5 and 10%. The SPA-LDA (Sucessive Projections Algorithm-Linear Discriminant Analysis) was efficient classifying correctly all the samples. The methods in study were efficient and promising allowing for phenotypic differentiation for progenitors and reciprocal generations of castor bean quickly, practice and non-destructively.

Keywords: Seed production, *Ricinus comunnis* L., VIS-NIR spectroscopy, reciprocal Generations, SIMCA, SPA-LDA

1. INTRODUÇÃO

A mamoneira se estabeleceu no Nordeste brasileiro como uma cultura voltada para o pequeno agricultor, baseado na disponibilidade de mão de obra familiar. Destacam-se na produção da espécie os estados da Bahia e do Ceará com mais de 80% da produção nacional. Notadamente o maior entrave para uma exploração mais racional e empresarial da cultura se deve entre outros fatores a baixa disponibilidade de sementes de cultivares adaptadas, produtivas, com elevado teor de óleo e tolerantes a pragas e doenças.

No Brasil o melhoramento genético da mamoneira já permitiu o desenvolvimento de cultivares mais produtivas, adaptadas a diversas regiões do país, apropriadas para diferentes tecnologias de colheita, resistente a algumas doenças e com alto teor de óleo na semente (FREIRE et al., 2001). No entanto, de acordo com Severino et al. (2012), entre as demandas atuais para o melhoramento genético da mamona, inclui-se a adaptação de genótipos à baixa altitude, o que permitirá a inclusão sustentável de muitos municípios onde o cultivo não é recomendado pelo risco de obtenção de baixas produtividades.

Para o sucesso do melhoramento da mamoneira é importante o conhecimento da natureza e da expressão dos caracteres agrônômicos da cultura, por intermédio da aplicação de técnicas eficientes de reconhecimento e avaliação desses materiais uma vez que permite a obtenção de ganhos genéticos satisfatórios (COSTA et al., 2006). As informações sobre a amplitude dos coeficientes de variabilidade e de herdabilidade nos permitem prognosticar o efeito da seleção e planejar os procedimentos do melhoramento (MOSHKIN, 1986).

Identificar corretamente os genótipos e suas respectivas progênes, tanto no estudo de heranças genéticas quanto na fenotipagem correta dos materiais tem adquirido maior importância a cada dia, pois é esperado que as sementes ao longo do processo de seleção e sucessivos cruzamentos vão adquirindo maior uniformidade. A busca por essa identificação e transferência destas características em programas de melhoramento que envolvam hibridações, resultam em práticas seletivas, ou ainda, na obtenção de híbridos F1 de expressivo vigor, ou heterose, decorrente do cruzamento entre dois genótipos de elevada capacidade combinatória e acentuada divergência genética entre si (MOREIRA, et al., 1996).

Figueiredo Neto et al. (2004) ressaltam a importância da caracterização e avaliação de cultivares para o estabelecimento de diferenças ou semelhanças entre acessos de germoplasma, bem como para estimular sua utilização e resgatar o desenvolvimento da cultura. A variabilidade em características botânicas e agrônômicas deve ser conhecida e conservada,

podendo se tornar fonte de genes para futuros programas de melhoramento. No entanto, a classificação correta e o conhecimento da variabilidade e diversidade dos acessos disponíveis são de extrema importância procedendo-se com base em diferenças na morfologia das plantas, nas moléculas de proteínas e de DNA (FERREIRA; GRATTAPAGLIA, 1998).

Os caracteres morfoagronômicos são eficientes para estudo da diversidade genética, pois proporcionam uma simplificação da quantificação da variação genética e possibilitam avaliar o desempenho dos genótipos no ambiente de crescimento (FUFA et al., 2005).

A uniformidade pode ser obtida pelo uso de cultivares de três tipos, a saber: híbridos, linhagens endogâmicas e clones. O uso de híbridos é um processo comum no caso de plantas alógamas, o uso de linhagens endogâmicas (variedades) é comum em plantas autógamas e o uso de clones é apropriado para espécies com propagação vegetativa (FHER, 1987; DESTRO; MONTALVÂN, 1999; NASS et al., 2001).

Os cruzamentos em plantas de mamona expressam a heterose ou vigor de híbrido, atingindo seu potencial máximo, posto que os híbridos de mamona são mais uniformes e produtivos.

A divergência genética entre um grupo de genitores tem sido, em muitos casos, avaliada com o objetivo de identificar as combinações híbridas de maior efeito heterótico, de tal forma que, em suas gerações segregantes, se tenha maior possibilidade de recuperação de genótipos superiores (FALCONER, 1989; CRUZ et al., 2004). A seleção dos genitores para o cruzamento deve-se aliar o bom desempenho destes com a divergência genética entre eles.

Os referidos estudos são de grande importância para o conhecimento da variabilidade genética das populações e, também, por possibilitar o monitoramento dos bancos de germoplasma, pois geram informações úteis para preservação e o uso de acessos (CRUZ; CARNEIRO, 2003). Essa variabilidade genética, espontânea ou criada, é o ponto de partida para qualquer programa de melhoramento, podendo ser alcançada por meio da escolha da dissimilaridade dos parentais, associado com média alta e variabilidade para as características de interesse. A população segregante, então, teria alta probabilidade de apresentar média elevada para as características de interesse, reduzindo, deste modo, o tempo e os custos gastos em programas de melhoramento (CRUZ, et al., 2004).

Para determinar quão distante geneticamente uma população é de outra, são utilizados métodos biométricos, os quais são analisados pela estatística multivariada permitindo unificar múltiplas informações de um conjunto de características extraídas das unidades experimentais, oferecendo ao melhorista maior oportunidade de escolha de genitores divergentes (SUDRÉ, et al., 2005).

De acordo com Ivoglo (2007), o estudo da divergência genética através de análises biométricas tem como objetivos: definição de população-base para seleção recorrente recíproca, identificação de genitores adequados à obtenção de híbridos com maior efeito heterótico e que também proporcionem maior segregação em recombinações, agrupamento de materiais genéticos similares para a formação de variedades sintéticas e caracterização da variabilidade de recursos genéticos em bancos de germoplasma “in situ” e “ex situ”.

Essa diferença genética entre indivíduos, tem sido avaliada por meio de técnicas biométricas, baseadas na quantificação da heterose, como nas análises dialélicas, e por processos preditivos (CRUZ et al., 2004); através da análise izoenzimática (BOUHARMONT et al., 1986; MONTAGNON et al., 1992; OLIVEIRA et al., 2002) e, pelo emprego de marcadores moleculares (OROZCO-CASTILLO et al., 1994; LASHERMES et al., 1996; CABRAL, 2001; ALVARENGA et al., 2005; SILVESTRINI et al., 2005).

Vários métodos preditivos podem ser utilizados no estudo da divergência genética, dentre eles estão a análise multivariada, por meio das medidas de dissimilaridade envolvendo a distância euclidiana e a distância generalizada de Mahalanobis; métodos de agrupamentos envolvendo os métodos hierárquicos, como UPGMA e do vizinho mais próximo e o método de otimização de Tocher; e técnicas de dispersão gráfica envolvendo análise por componentes principais e por variáveis canônicas (CRUZ et al., 2004). A escolha do método mais adequado tem sido determinada pela precisão desejada, pela facilidade da análise e ainda pela forma como os dados foram obtidos (CRUZ; CARNEIRO, 2003; CRUZ et al., 2004).

Nesse sentido a espectrometria no infravermelho próximo vem sendo empregado como método preditivo na identificação de genótipos baseados nas alterações químicas existentes na amostra.

Cui et al., (2012) desenvolvendo estudo para discriminação rápida e precisa de três genótipos de algodão sendo os pais, transgênicos e híbridos de pais transgênicos por meio de análise de NIR em folhas e sementes, obtiveram modelo de classificação total de precisão de 100% e 97,6% para as sementes e folhas, respectivamente, comprovando a eficiência do uso do método.

Han et al, (2014) avaliando populações segregantes F2 e F3 a partir de linhagens de soja com elevados teores de ácido oleico desenvolveram a partir de 600 sementes individuais equações de calibração para identificar sementes com mais de 70% de ácido oleico utilizando a espectroscopia NIR. Os espectros foram coletados entre 400 e 2.500nm a 2nm com intervalos

variáveis. Os resultados mostraram que o NIR previu a alta do teor ácido oleico nas populações avaliadas e constitui um método satisfatório para predição

Poucos são os trabalhos utilizando a espectroscopia VIS-NIR em sementes intactas e em sucessivas gerações, desta forma, a tecnologia se torna uma nova perspectiva em identificações de diferentes sementes de mamona de forma rápida, prática, barata e não destrutiva

Com isso, o objetivo deste estudo é identificar marcadores espectrais na região VIS-NIR que promovam diferenciação entre genitores e sua descendência F1, F2, F3 e F4.

2. MATERIAL E METODOS

2.1. Aquisição das Amostras

Foram utilizados dois parentais sendo CNPAM 2012-137 e CNPAM 93-168 e quatro gerações recíprocas F1, F2, F3 e F4 autofecundadas por dois anos e cedidas pelo programa de melhoramento da Embrapa Algodão para este estudo (**Figura 25**).

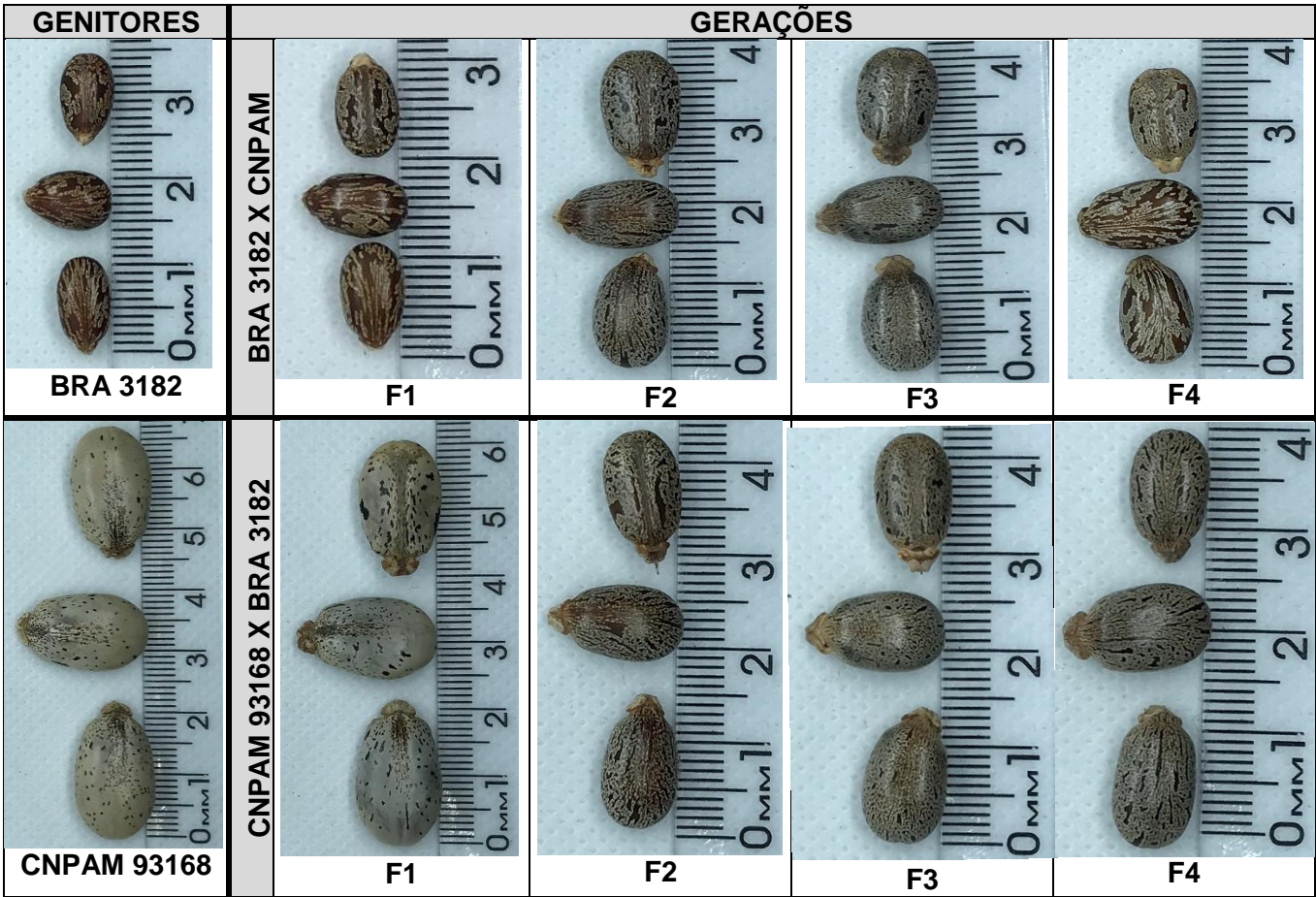


Figura 25- Sementes de mamona de acordo com seus parentais e suas gerações recíprocas.

Para cada material foram empregadas vinte sementes, sendo acondicionadas a uma temperatura de 21°C e umidade relativa de 70% para realização das avaliações. A quantidade limitada de amostra é um dos fatores sempre recorrentes em melhoramento vegetal, neste caso, em particular, o número máximo de amostras (genótipos) trabalhados foi de 20 sementes.

2.2. Instrumentação

As medidas de reflectância difusa foram obtidas em um espectrômetro VIS-NIR modelo XDS Rapid Content™ Analyser (Foss Analytical, Hogans, Sweden) conforme ilustrado na **figura 26 (A)**. Na **figura 26 (B)** é possível observar no detalhe o ambiente de leitura da amostra.



Figura 26- (A) Espectrofotômetro VIS-NIR; (B) Detalhe ambiente de leitura da amostra

Na **Figura 27 (A)** observa-se a célula de quartzo circular de 3 cm de diâmetro usada para posicionar a amostra a ser analisada. Para bloquear a radiação espúria do ambiente, foram usadas tampas reflexivas na célula de amostragem, **Figura 27 (B)**.

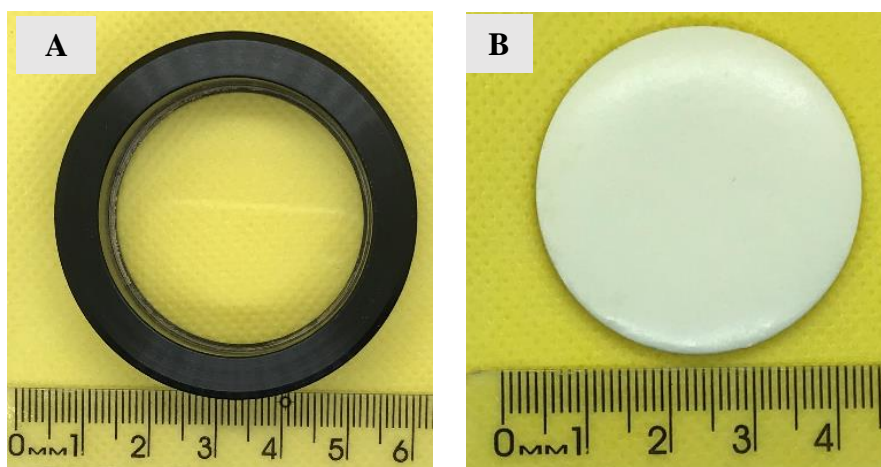


Figura 27- (A) Célula de quartzo; 36 (B) tampas reflexivas para a célula de quartzo.

Inicialmente com o intuito do registro do sinal de base utilizou-se um padrão de reflectância conforme observado na **figura 28**.



Figura 28- Padrão de reflectância

2.3. Aquisição dos espectros NIR

Os espectros de reflectância foram obtidos diretamente sem nenhum tratamento químico das sementes.

As amostras de sementes individuais sempre foram dispostas na célula, da mesma maneira para assegurar a uniformidade das medidas. Cada espectro foi obtido a partir de 32 varreduras na faixa de 400 a 2500 nm em intervalos de 0,5 nm.

2.4. Programas Computacionais

O pré-processamento dos espectros originais e a aplicação das técnicas de reconhecimento de padrões não supervisionado (Análise de Componentes Principais-PCA) e supervisionado (SIMCA) foram realizados utilizando-se o programa Unscrambler® 8. A aplicação do algoritmo Kennard-Stone utilizado para seleção de amostras e a modelagem SPA-LDA foram realizadas em ambiente Matlab R2008a.

2.5. Tratamento Quimiométrico dos Dados

2.5.1. Pré-processamento

A região espectral de 400 a 750nm foi selecionada a priori como a região de trabalho para classificação de sementes de mamoneira que corresponde a região do visível VIS-NIR.

As técnicas de suavização Savitzky-Golay, correção multiplicativa de sinais (MSC), correção de linha de base e derivação, foram avaliadas no pré-processamento dos espectros.

2.5.2. Reconhecimento de Padrões

Realizou-se uma análise exploratória utilizando-se a PCA com o objetivo de observar a formação de agrupamentos.

O algoritmo Kennard-Stone foi aplicado separadamente aos espectros de cada cultivar de mamona com a finalidade de dividir as amostras em conjuntos de validação cruzada e teste, conforme a **tabela 9** e também em conjunto de treinamento, validação e teste (**Tabela 10**). Esses conjuntos foram utilizados na análise de classificação SIMCA e na modelagem SPA-LDA.

Tabela 9. Número de amostras dos conjuntos de validação cruzada e teste, selecionadas pelo algoritmo KS, para os progenitores e os cruzamentos recíprocos.

Cultivar/	Conjuntos		Total
	Validação cruzada	Teste	
CNPAM 2012-137	14	6	20
CNPAM 93-168	14	6	20
F1 CNPAM 2012-137 x CNPAM 93-168	14	6	20
F1 CNPAM 93-168 x CNPAM 2012-137	14	6	20
F2 CNPAM 2012-137 x CNPAM 93-168	14	6	20
F2 CNPAM 93-168 x CNPAM 2012-137	14	6	20
F3 CNPAM 2012-137 x CNPAM 93-168	14	6	20
F3 CNPAM 93-168 x CNPAM 2012-137	14	6	20
F4 CNPAM 2012-137 x CNPAM 93-168	14	6	20
F4 CNPAM 93-168 x CNPAM 2012-137	14	6	20

Tabela 10. Número de amostras dos conjuntos de treinamento, validação e teste, selecionadas pelo algoritmo KS, para os progenitores e os cruzamentos recíprocos.

Cultivar/	Conjuntos			Total
	Treinamento	Validação	Teste	
CNPAM 2012-137	10	4	6	20
CNPAM 93-168	10	4	6	20
F1 CNPAM 2012-137 x CNPAM 93-168	10	4	6	20
F1 CNPAM 93-168 x CNPAM 2012-137	10	4	6	20
F2 CNPAM 2012-137 x CNPAM 93-168	10	4	6	20
F2 CNPAM 93-168 x CNPAM 2012-137	10	4	6	20
F3 CNPAM 2012-137 x CNPAM 93-168	10	4	6	20
F3 CNPAM 93-168 x CNPAM 2012-137	10	4	6	20
F4 CNPAM 2012-137 x CNPAM 93-168	10	4	6	20
F4 CNPAM 93-168 x CNPAM 2012-137	10	4	6	20

3. RESULTADOS E DISCUSSÃO

3.1. Espectros NIR

Na **figura 29** são observados os espectros originais e na **figura 30** os espectros derivativos de 100 amostras das diferentes sementes dos progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500nm.

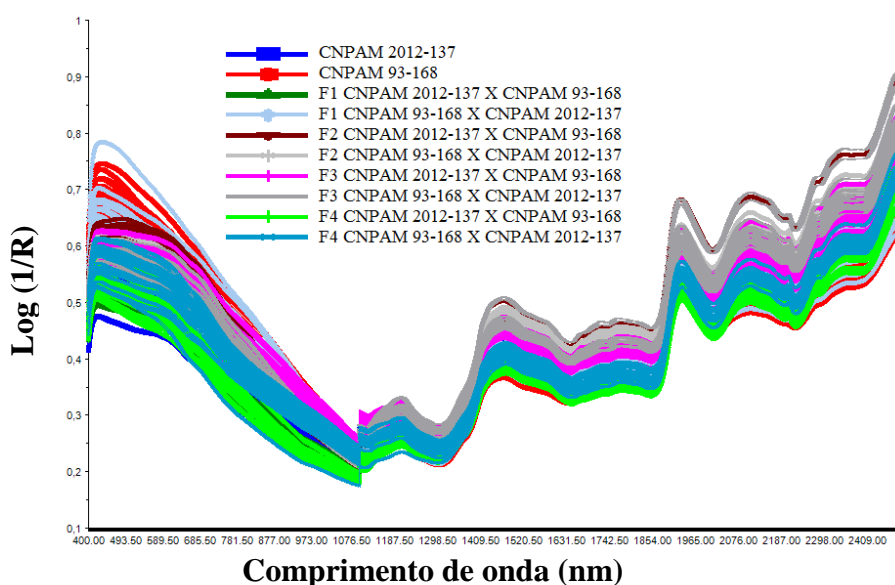


Figura 29- Espectros originais NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500nm

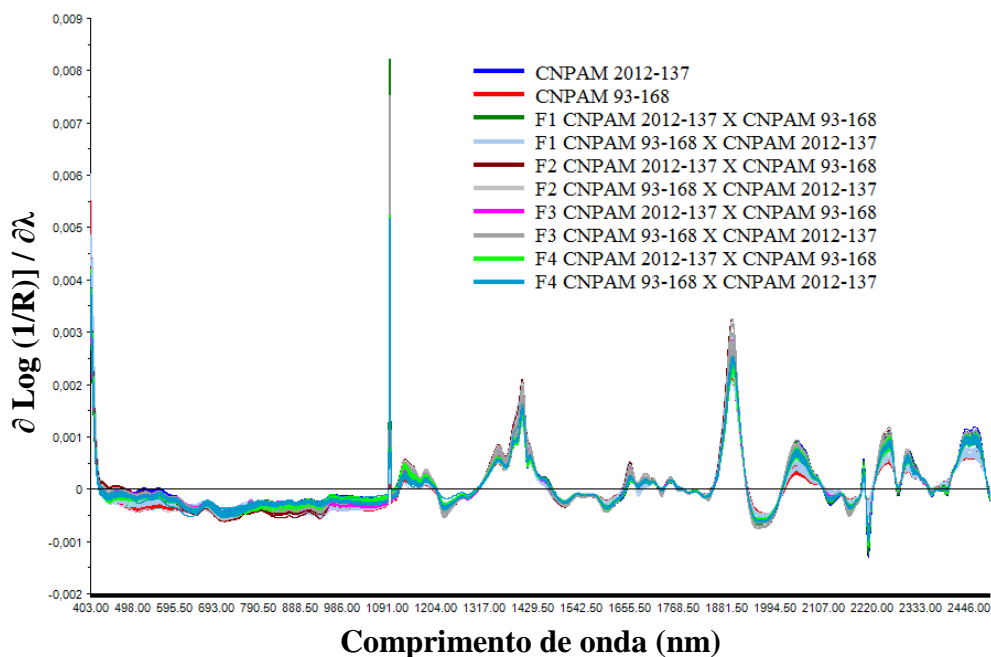


Figura 30- Espectros derivativos NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500nm.

Devido a complexidade da informação foram calculados os espectros médios originais (**figura 31**) e derivativos (**figura 32**) de reflectância difusa das sementes de mamona, de diferentes sementes dos progenitores e os cruzamentos recíprocos obtidos entre 400 e 2500 nm.

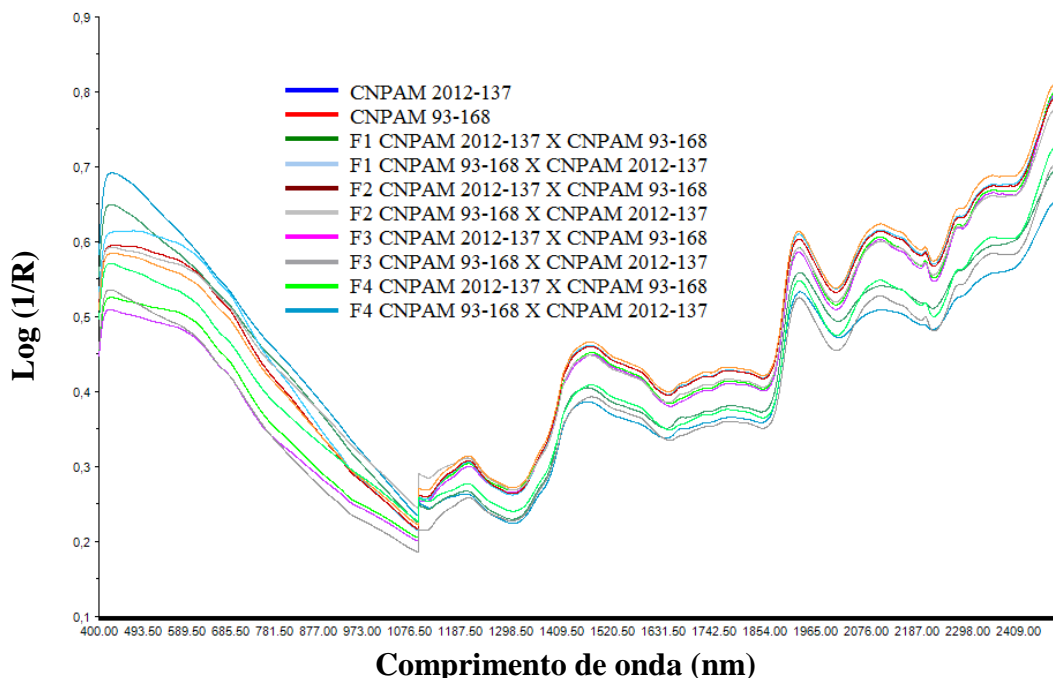


Figura 31- Espectros médios originais NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e cruzamentos recíprocos obtidos entre 400 e 2500nm.

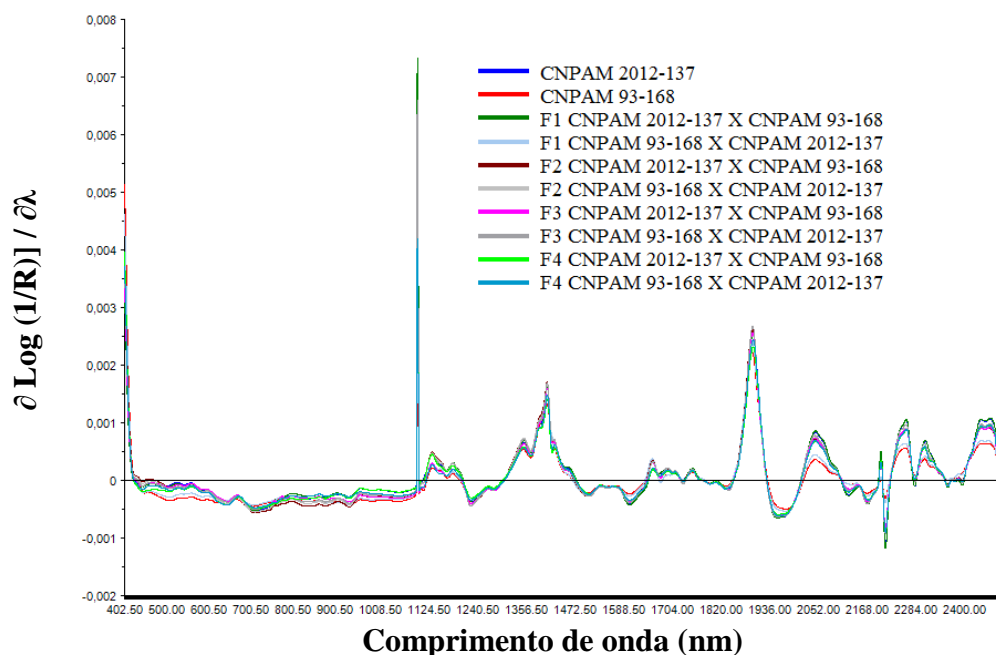


Figura 32- Espectros médios derivativos NIR de reflectância difusa das sementes de mamona, de diferentes progenitores e cruzamentos recíprocos obtidos entre 400 e 2500nm.

Em seguida como pode ser observado na **figura 33**, foi selecionada apenas a faixa do visível (400-750nm) que corresponde a região de trabalho, em seus espectros originais e na **figura 34** os espectros derivativos.

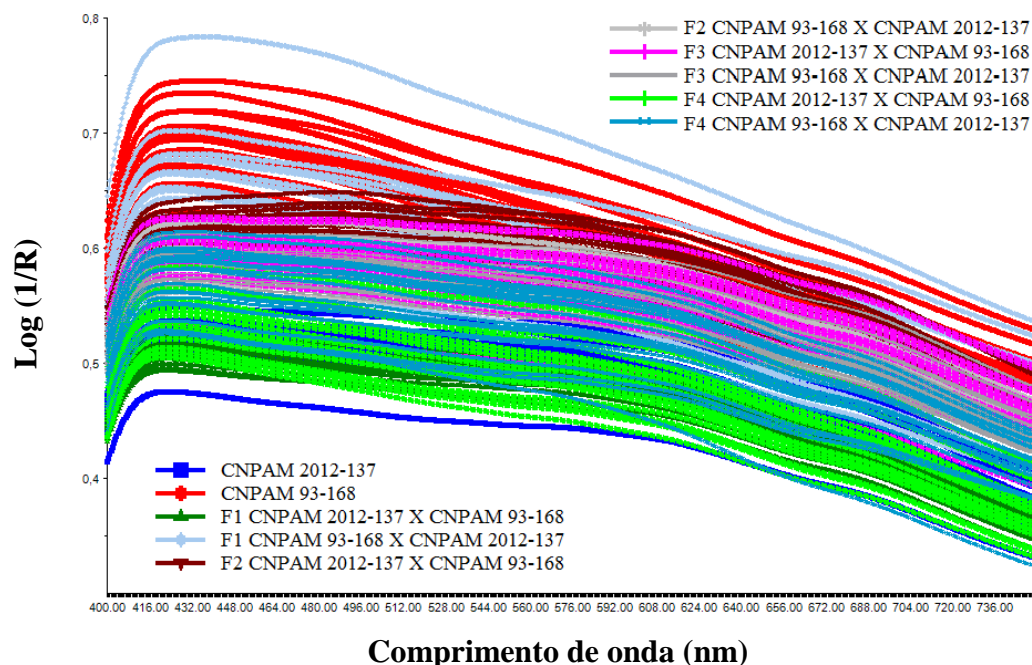


Figura 33- Espectros originais VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.

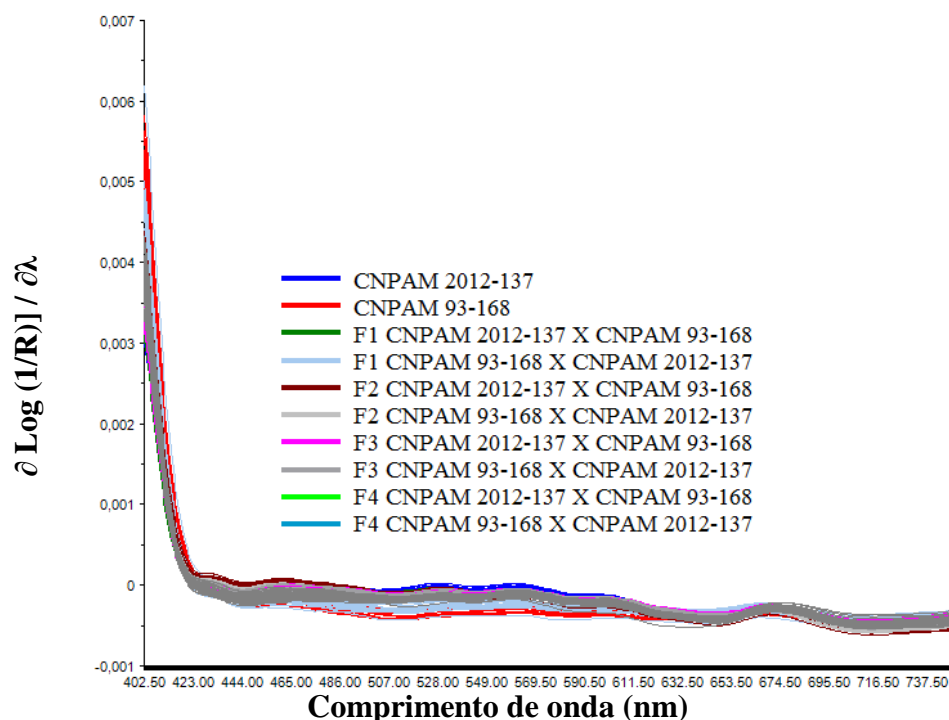


Figura 34- Espectros derivativos VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.

Buscando ainda uma melhor visualização em relação aos genótipos, foram calculados os espectros médios na faixa do VIS-NIR originais (**Figura 35**) e derivativos (**Figura 36**). Nesses espectros não há ruído instrumental evidente. Contudo, uma alteração do perfil de linha de base pode ser observada. Esta foi corrigida empregando-se a primeira derivada, com o filtro de Savitzky-Golay, polinômio de segunda ordem e uma janela de 11 pontos.

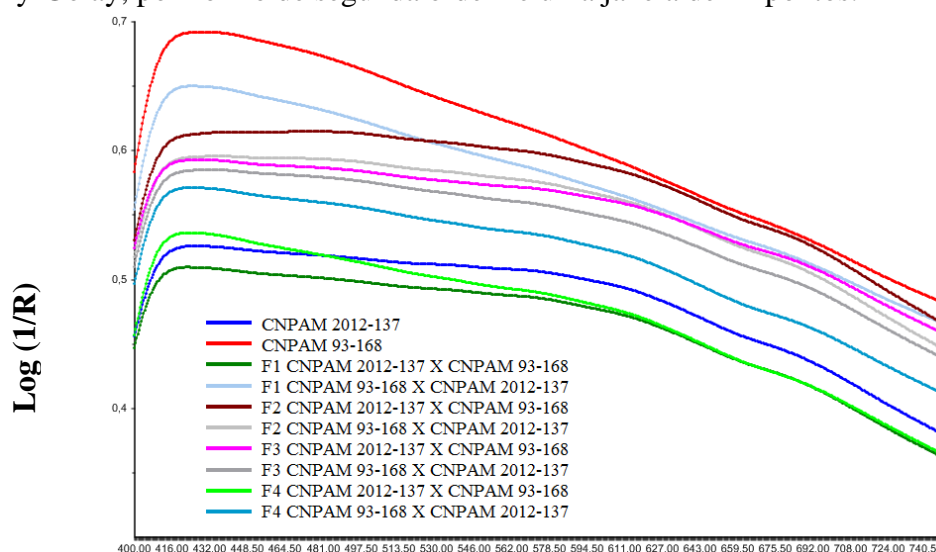


Figura 35- Espectros médios originais VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.

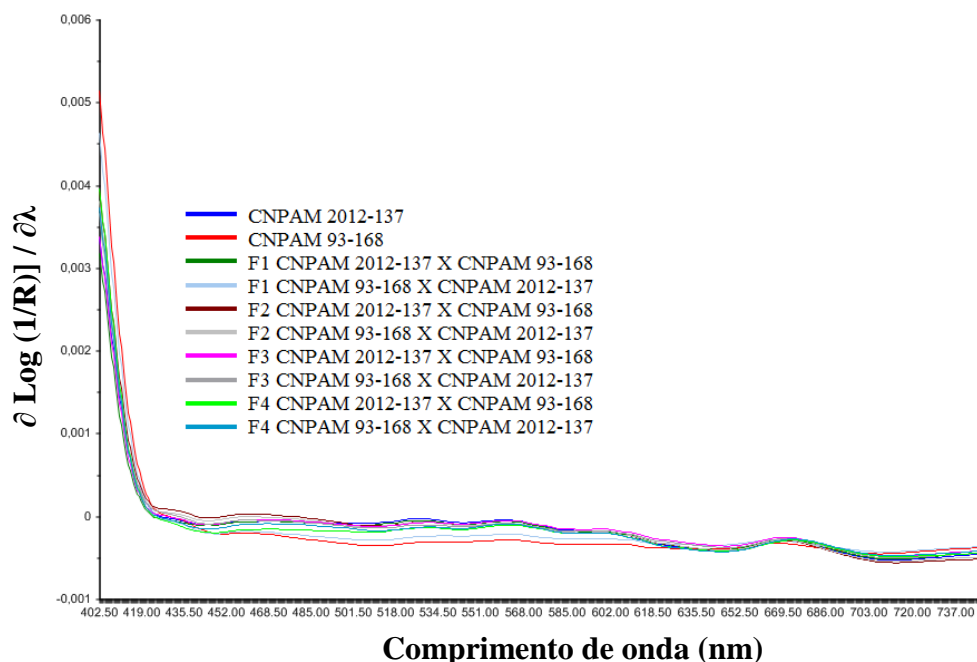


Figura 36- Espectros médios derivativos VIS-NIR de reflectância difusa das sementes de mamona de diferentes progenitores e cruzamentos recíprocos, na faixa de 400-750nm.

Observando os espectros derivativos das 60 amostras de sementes de mamoneira visualizados na **figura 36** na faixa VIS-NIR, é possível observar que mesmo após a correção do incremento de linha de base com o procedimento empregado, algumas das amostras analisadas em faixas específicas, possuem perfis espectrais semelhantes e sobrepostos, o que dificulta a melhor visualização.

Mesmo havendo a possibilidade de separação visual dos progenitores, a complexidade do sinal obtido e a semelhança existente na composição química das sementes impossibilitam a distinção visual de alguns dos cruzamentos recíprocos. Neste contexto, torna-se necessário o uso de ferramentas quimiométricas.

3.2. Análise Exploratória dos Dados

Na **figura 37** observa-se o gráfico dos escores em **A** (PC1 vs PC2) e em **B** (PC1 vs PC3) e na **47** o gráfico dos pesos (PC1) como resultado da aplicação da PCA a todos os espectros pré-processados

A discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1, explicando 84% da variância encontrada. Com a finalidade de verificar os principais

comprimentos de onda, responsáveis por tal efeito, foi examinado, o gráfico de pesos de PC1 (Figura 38).

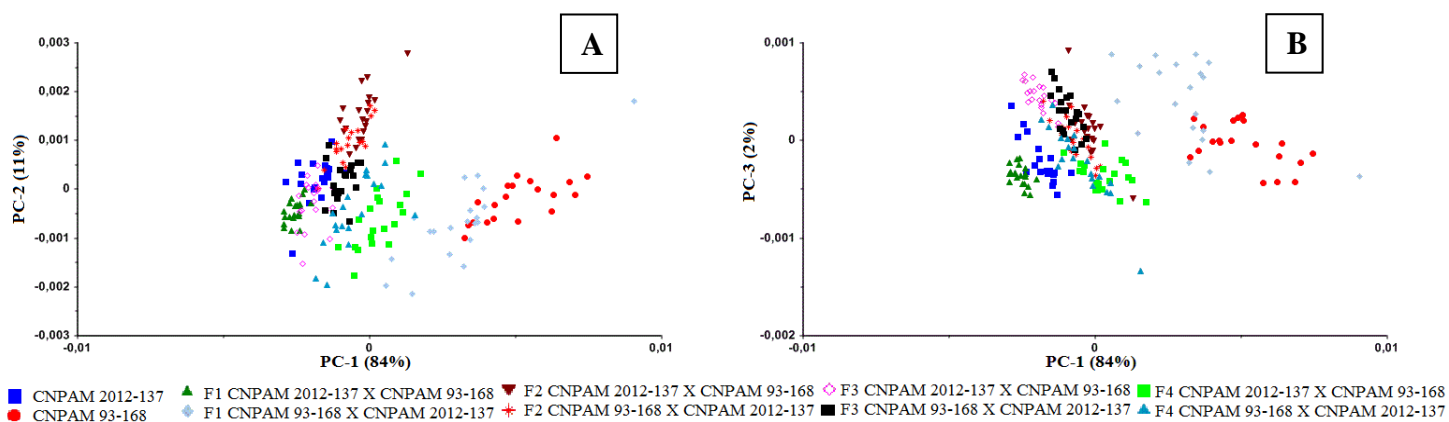


Figura 37- Gráfico dos escores em **A** (PC1 vs PC2) e em **B** (PC1 vs PC3) para o conjunto de 200 amostras como resultado da aplicação da PCA a todos os espectros pré-processados.

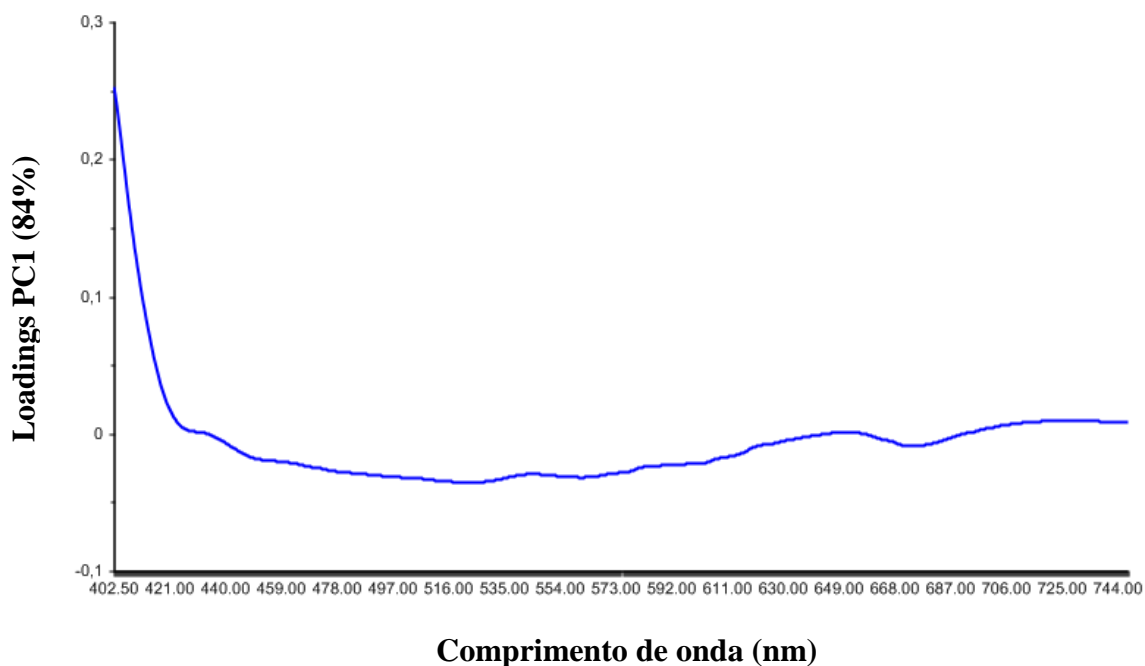


Figura 38- Gráfico dos pesos (PC1) para o conjunto de 200 amostras como resultado da aplicação da PCA a todos os espectros pré-processados.

Na **figura 39** observa-se o gráfico dos escores e na **figura 40** o gráfico dos pesos como resultado da aplicação da PCA (PC1 versus PC2) aos espectros pré-processados dos genitores CNPAM 2012-137e CNPAM 93-168.

Com base no gráfico de escores há uma nítida separação entre as amostras dos genitores CNPAM 2012-137e CNPAM 93-168 em PC1. Observa-se, ainda, no gráfico dos escores, que não existem amostras isoladas e, portanto, as 20 amostras foram utilizadas.

Com a finalidade de verificar os principais comprimentos de onda, responsáveis por tal efeito, foi examinado, o gráfico de pesos de PC1.

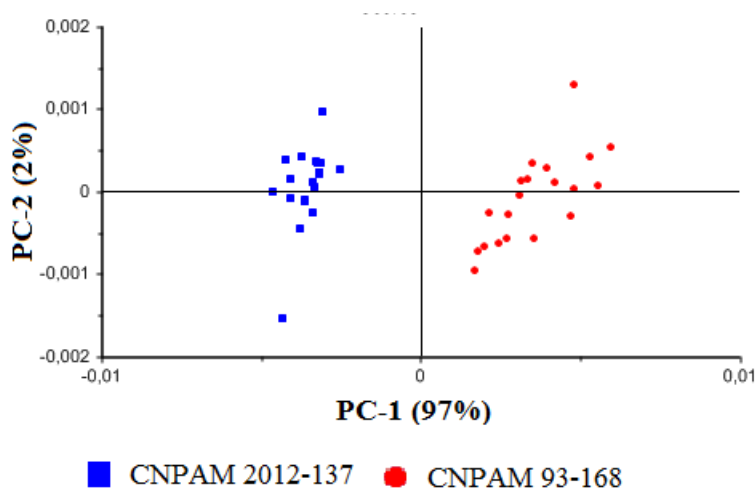


Figura 39- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos genitores CNPAM 2012-137e CNPAM 93168

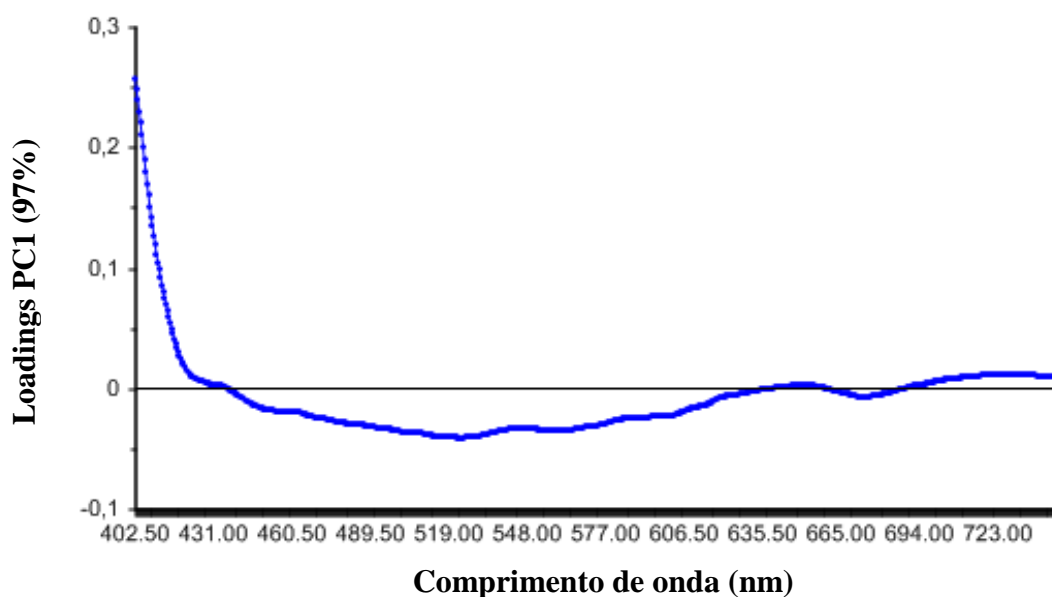


Figura 40- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos genitores CNPAM 2012-137 e CNPAM 93-168

Na **figura 41** observa-se o gráfico dos escores na **figura 42** o gráfico dos pesos como resultado da aplicação da PCA (PC1 versus PC2) aos espectros pré-processados dos cruzamentos recíprocos F1.

Baseado no gráfico de escores visualiza-se a separação entre as amostras dos cruzamentos recíprocos F1 em PC1. A discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1. Com a finalidade de verificar os principais comprimentos de onda, responsáveis por tal efeito, foi examinado, o gráfico de pesos de PC1.

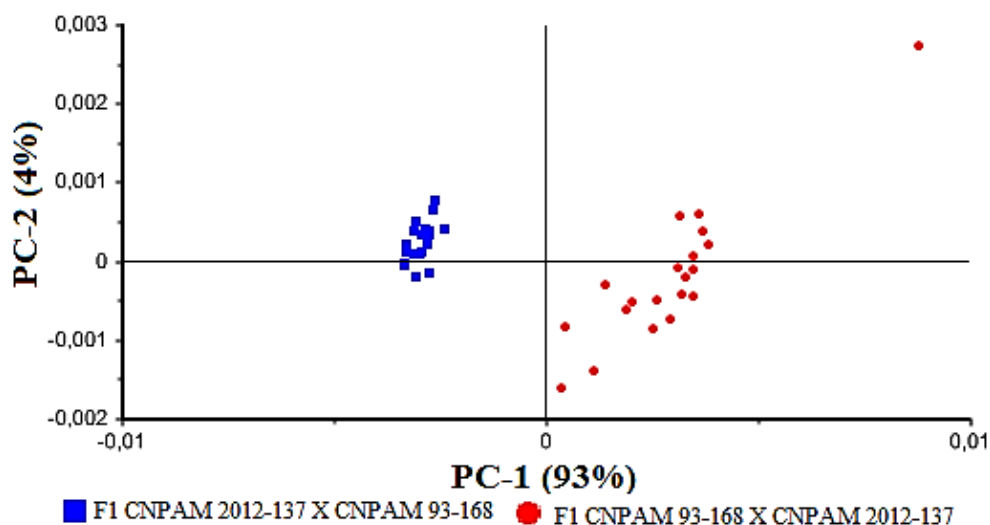


Figura 41- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F1.

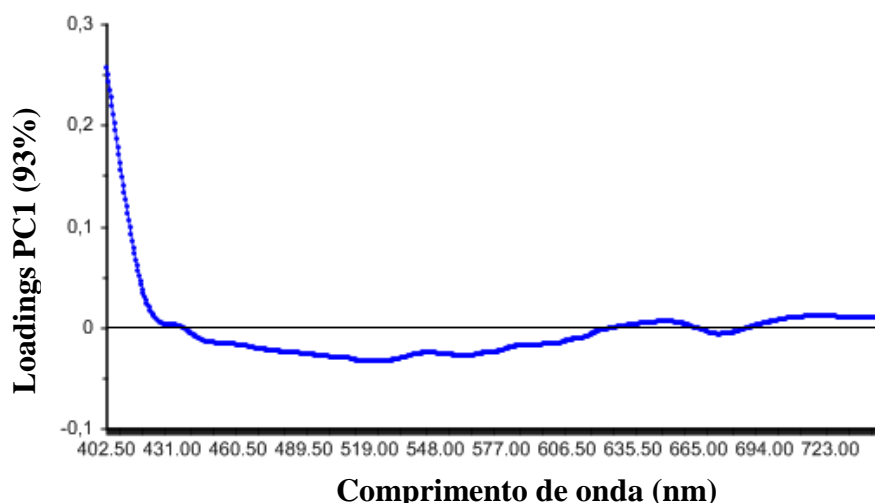


Figura 42- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F1.

Na **figura 43** observa-se o gráfico dos escores como resultado da aplicação da PCA (PC1 versus PC2) aos espectros pré-processados dos cruzamentos recíprocos F2. Visualiza-se a

separação entre as amostras dos cruzamentos recíprocos F2 em PC1, contudo, também ocorre uma sobreposição entre as classes o que, possivelmente, pode vir a comprometer o desempenho dos modelos de classificação.

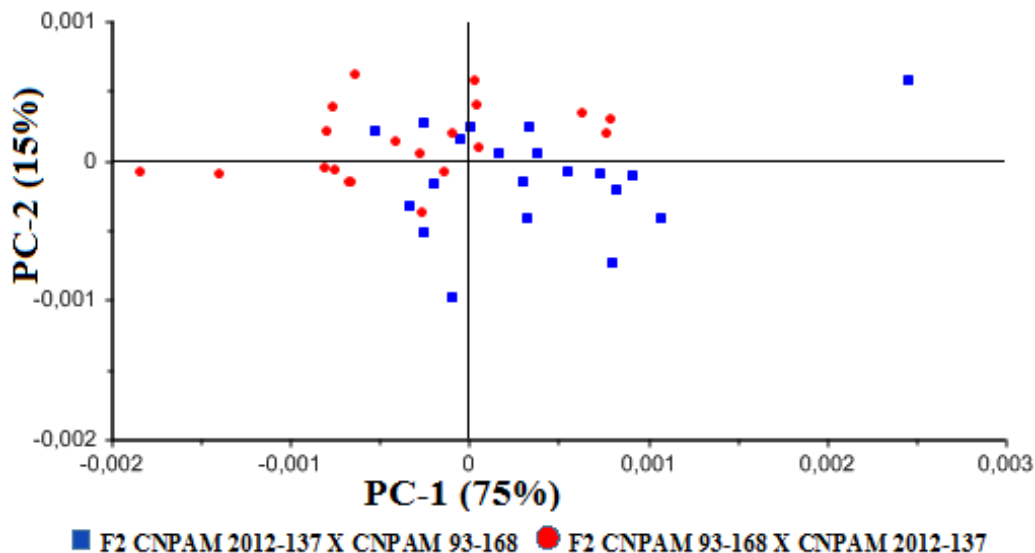


Figura 43- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F2.

Na **figura 44**, visualizamos o gráfico de pesos em PC1, estando correlacionada a discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1, sendo possível verificar os principais comprimentos de onda, responsáveis por tal efeito.

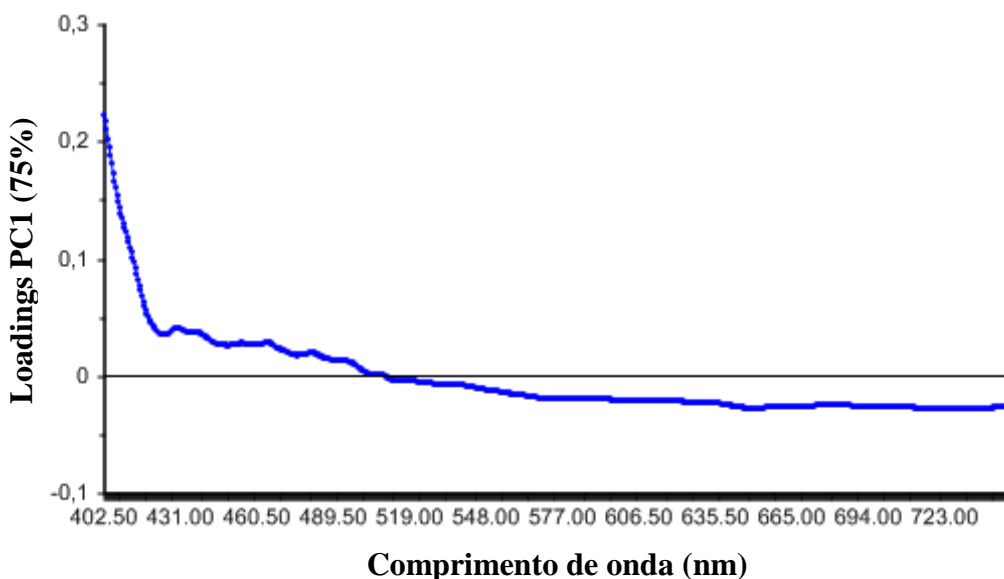


Figura 44- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F2.

Na **figura 45** observa-se o gráfico dos escores (PC1 versus PC2) e na **figura 46** o gráfico dos pesos (PC1) como resultado da aplicação da PCA aos espectros pré-processados dos cruzamentos recíprocos F3. Baseado no gráfico de escores observa-se a separação entre as amostras dos cruzamentos recíprocos F3 em PC1, contudo, também ocorre uma sobreposição entre as classes o que, possivelmente, pode vir a comprometer o desempenho dos modelos de classificação.

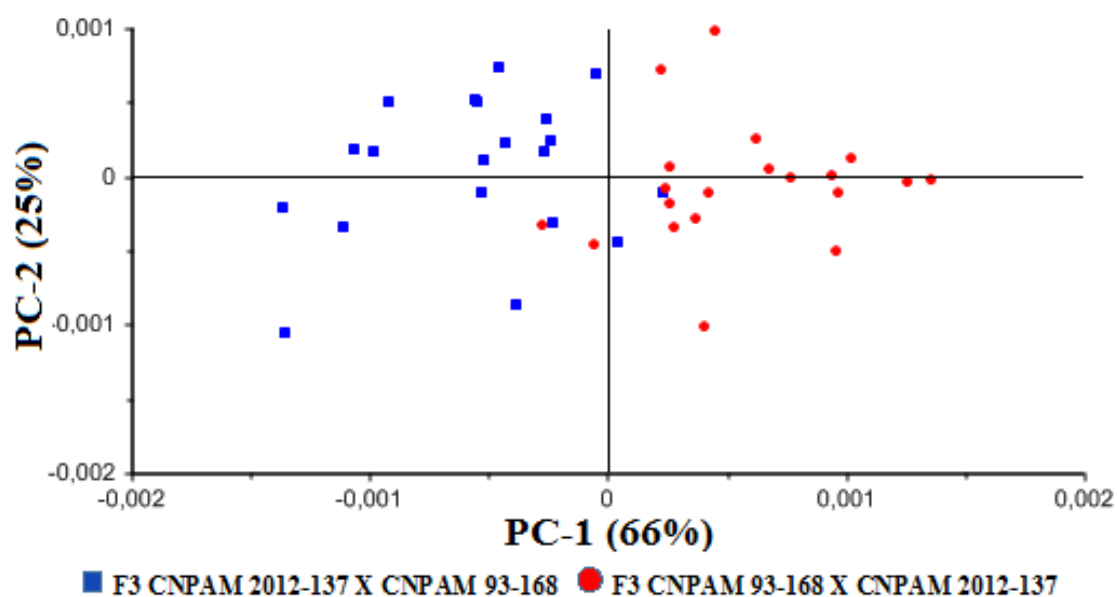


Figura 45- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F3.

A discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1. Com a finalidade de verificar os principais comprimentos de onda, responsáveis por tal efeito, foi examinado, o gráfico de pesos de PC1.

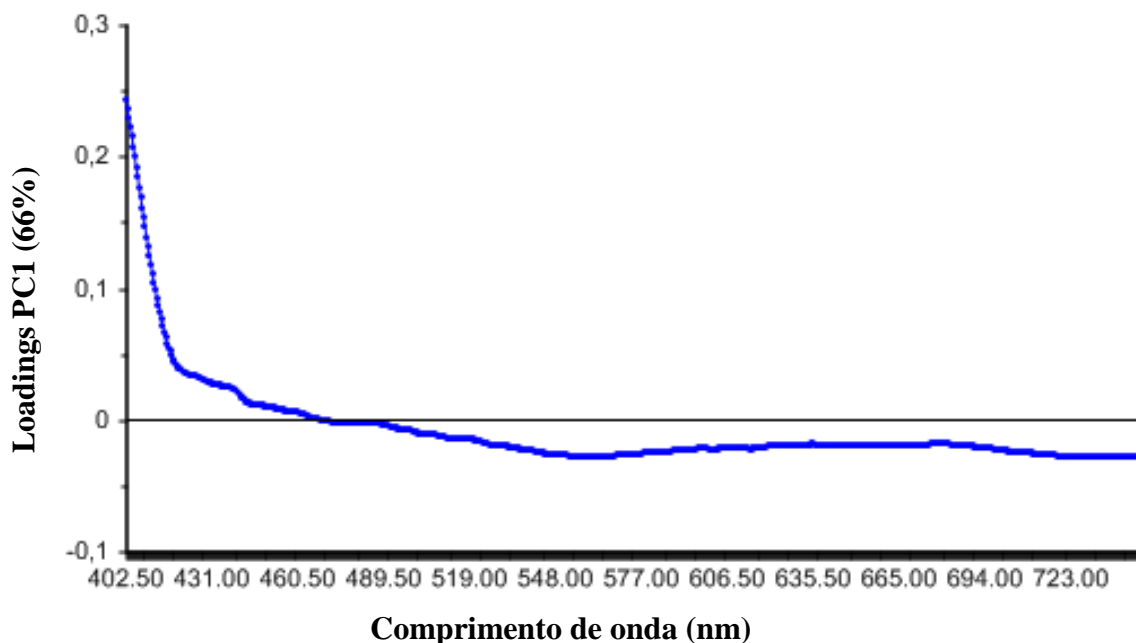


Figura 46- Gráfico dos pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F3.

Na **figura 47** observa-se o gráfico dos escores e na **figura 48** o gráfico dos pesos como resultado da aplicação da PCA (PC1 versus PC2) aos espectros pré-processados dos cruzamentos F4.

Com base no gráfico de escores visualizamos a separação entre as amostras dos cruzamentos recíprocos F4 em PC1, contudo, também ocorre uma sobreposição entre as classes o que deve ser melhor avaliado, uma vez que sobreposições podem vir a comprometer o desempenho dos modelos de classificação.

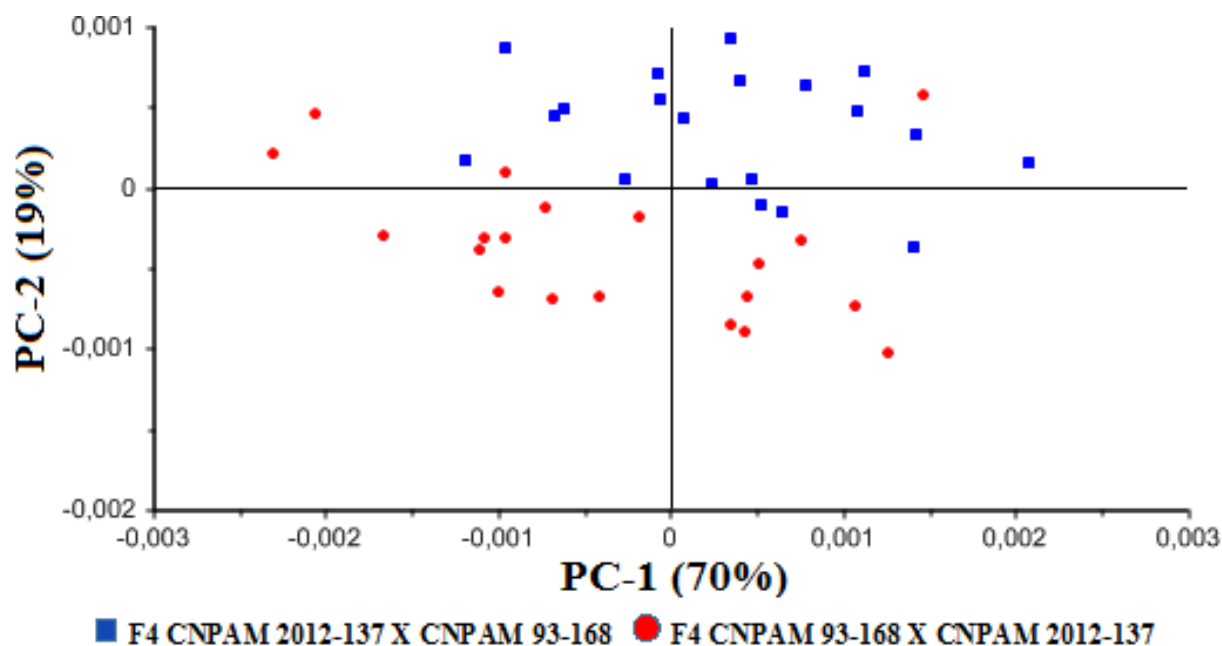


Figura 47- Gráfico dos escores (PC1 vs PC2) para o conjunto de 20 amostras dos cruzamentos recíprocos F4.

A discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1. Com a finalidade de verificar os principais comprimentos de onda, responsáveis por tal efeito, foi examinado, o gráfico de pesos de PC1.

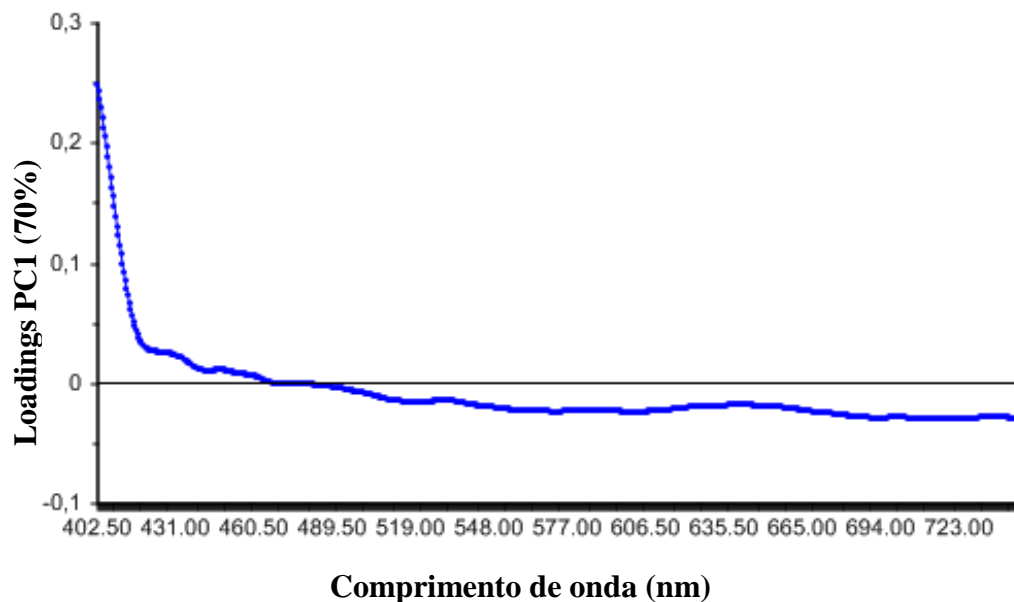


Figura 48- Gráfico de pesos (PC1) para o conjunto de 20 amostras dos cruzamentos recíprocos F4.

Foi efetuada a classificação isolada entre cada genitor e sua descendência direta com o intuito de testar o modelo empregado na separação das amostras.

Na **figura 49** observa-se o gráfico dos escores (PC1 versus PC2) como resultado da aplicação da PCA aos espectros pré-processados do genitor CNPAM 2012-137e seus cruzamentos F1, F2, F3 e F4.

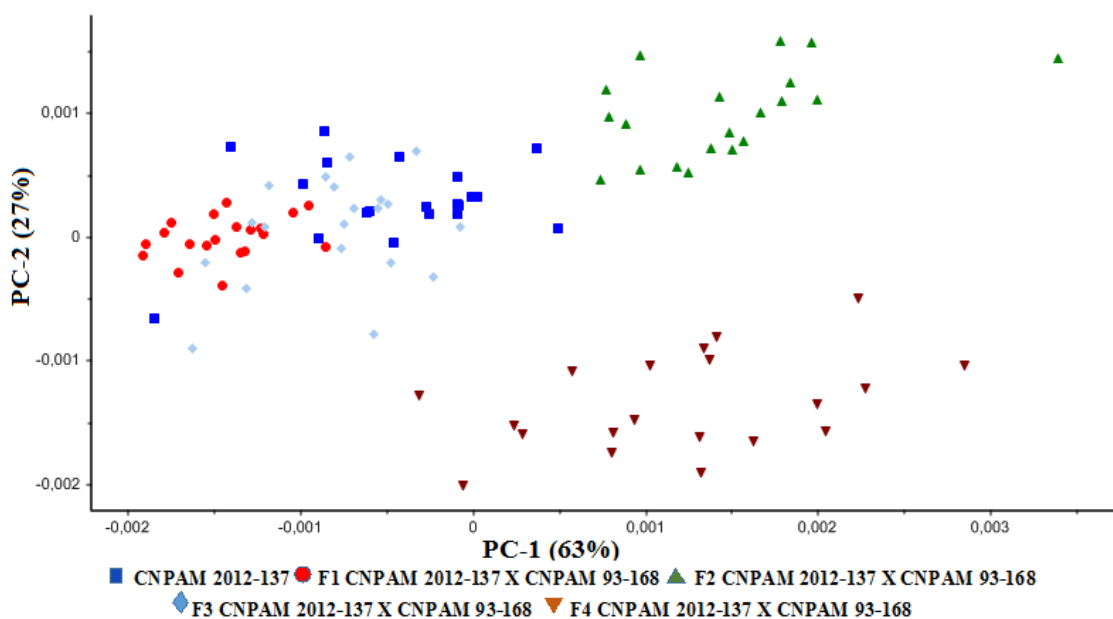


Figura 49- Gráfico dos escores (PC1 vs PC2) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 2012-137e seus cruzamentos F1, F2, F3 e F4.

Com base no gráfico de escores visualiza-se a separação entre o genitor e sua descendência F1, F2 e F3, sendo observadas poucas áreas de sobreposição, contudo, ocorre considerável separação em relação a F4 o que, possivelmente, pode ter acontecido devido ao processo de melhoramento ter conduzido a uma seleção com características não encontradas nos demais agrupamentos. Esses resultados nos permitem correlacionar com a legislação atual de sementes em relação as classes certificadas: Básica, C1 e C2.

A discriminação entre as amostras de sementes de mamona ocorre praticamente na PC1. Sendo analisados os principais comprimentos de onda, responsáveis por tal efeito o gráfico de pesos de PC1 (**Figura 50**).

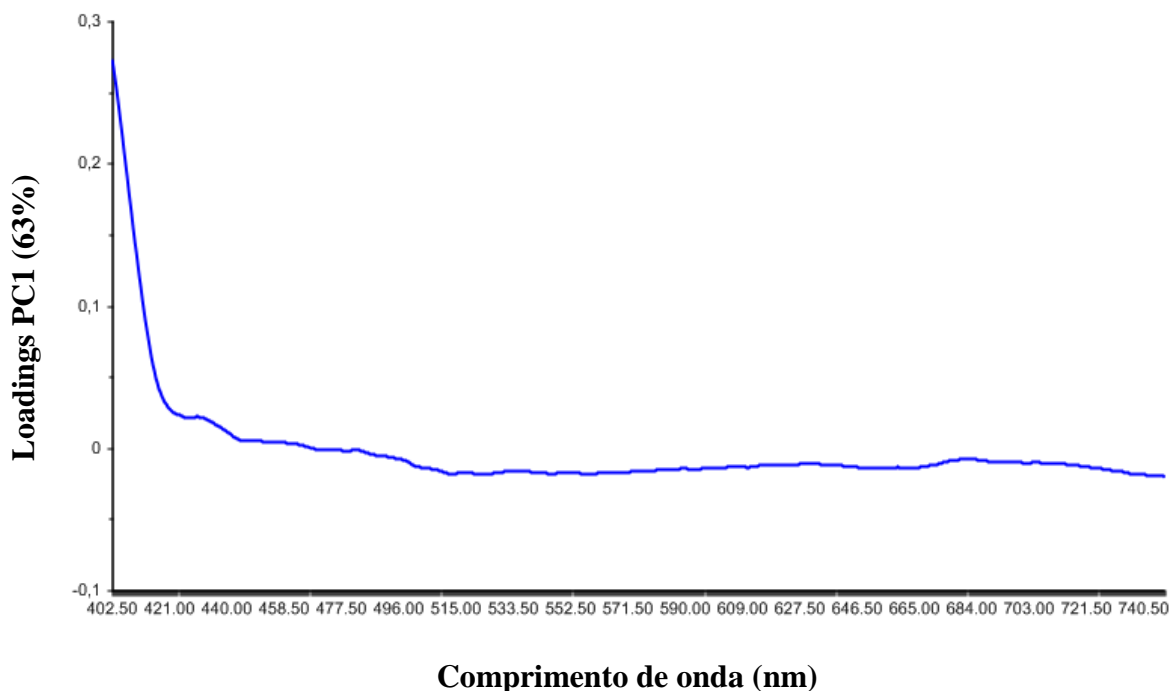


Figura 50- Gráfico dos pesos (PC1) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 2012-137e seus cruzamentos F1, F2, F3 e F4.

Na **figura 51** observa-se o gráfico dos escores (PC1 versus PC2) como resultado da aplicação da PCA aos espectros pré-processados do genitor CNPAM 93-168 e seus cruzamentos F1, F2, F3 e F4. É visualizada separação entre o genitor e sua descendência F1, F2, F3 e F4 sendo observadas áreas de sobreposição entre o parental e a F1. Visualiza-se ainda, poucas áreas de sobreposição em relação as gerações F1, F2, F3 e F4. As sobreposições possivelmente, podem vir a comprometer o desempenho dos modelos de classificação, o que deve ser melhor avaliado.

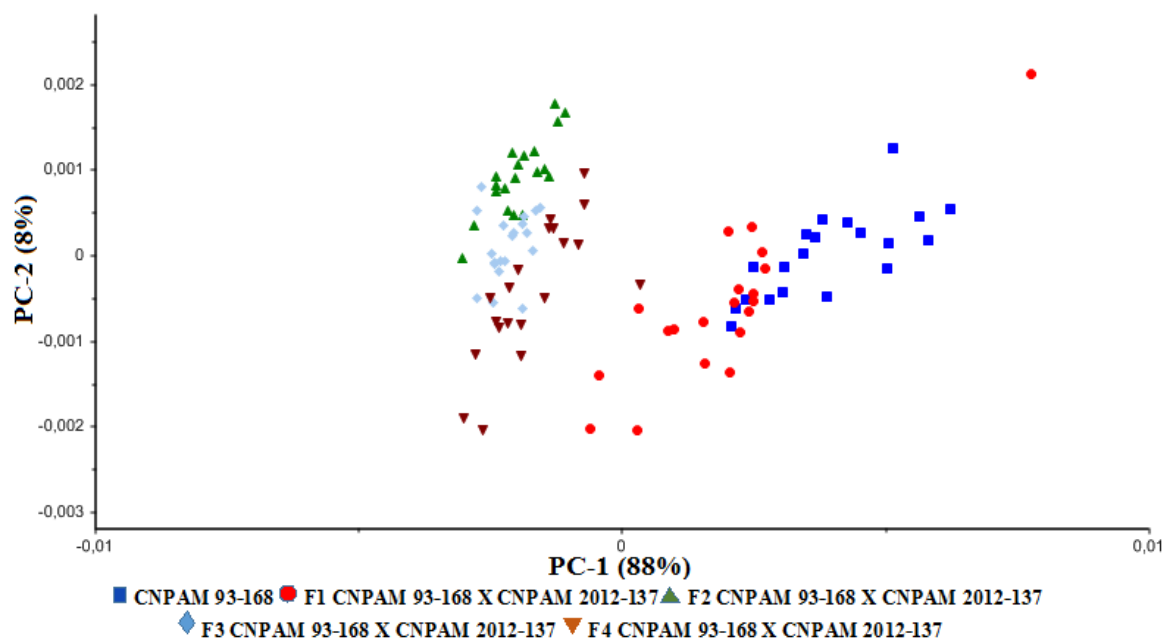


Figura 51- Gráfico dos escores (PC1 vs PC2) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 93-168 e seus cruzamentos F1, F2, F3 e F4.

Analisou-se os principais comprimentos de onda, responsáveis por tal efeito no gráfico de pesos de PC1 (**Figura 52**)

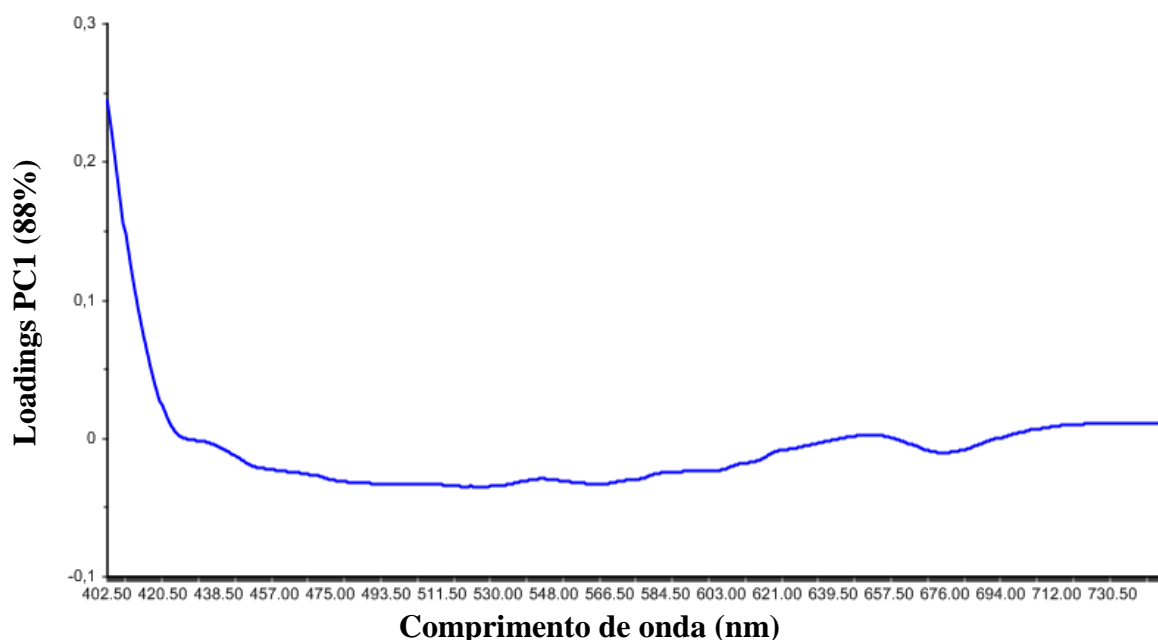


Figura 52- Gráfico dos pesos (PC1) para o conjunto de 100 amostras dos cruzamentos do genitor CNPAM 93-168 e seus cruzamentos F1, F2, F3 e F4.

3.3. *Reconhecimento de Padrões Supervisionados*

3.3.1. *Construção e Validação dos Modelos SIMCA*

Após a construção das PCAs baseadas em conjuntos completos das amostras dos progenitores e todos os cruzamentos recíprocos, separadamente, validados empregou-se um conjunto externo de amostras. Ficou evidenciado que os grupos formados, possuem homogeneidade de comportamento, demonstrando a eficiência da seleção das amostras de validação realizadas com o algoritmo KS.

No total, quatro PCs foram selecionadas para todas as classes, estas explicam mais de 90% do modelo para todas as classes avaliadas. O número ótimo de PCs na PCA será usado na classificação SIMCA.

Na **tabela 10** são ilustrados os erros de classificação do conjunto de validação com o objetivo de avaliar o desempenho dos modelos construídos.

Tabela 11. Erros de classificação ao nível de 1, 5, 10 e 25% de confiança para teste F.

AMOSTRAS DE PREVISÃO											
Modelos	%	CNPAM 2012-137	CNPAM 93- 168	F1- CNPAM 2012-137	F1- CNPAM 93-168	F2- CNPAM 2012-137	F2- CNPAM 93-168	F3- CNPAM 2012-137	F3- CNPAM 93-168	F4- CNPAM 2012-137	F4- CNPAM 93-168
CNPAM 2012-137	1	1	0	4	0	0	0	0	0	0	0
	5	1	0	4	0	0	0	0	0	0	0
	10	1	0	4	0	0	0	0	0	0	0
	25	4	0	1	0	0	0	0	0	0	0
CNPAM 93-168	1	0	0	0	6	0	0	0	0	0	0
	5	0	0	0	6	0	0	0	0	0	0
	10	0	0	0	6	0	0	0	0	0	0
	25	0	0	0	5	0	0	0	0	0	0
F1- CNPAM 2012- 137	1	0	0	0	0	0	0	0	0	0	0
	5	4	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0
	25	0	0	0	0	0	0	0	0	0	0
F1- CNPAM 93- 168	1	0	6	0	0	0	0	0	0	0	0
	5	0	6	0	0	0	0	0	0	0	
	10	0	6	0	0	0	0	0	0	0	0
	25	0	6	0	1	0	0	0	0	0	0
F2- CNPAM 2012- 137	1	0	0	0	0	2	6	0	0	0	0
	5	0	0	0	0	2	6	0	3	0	0

	10	0	0	0	0	2	6	0	2	0	0
	25	0	0	0	0	2	5	0	1	0	0
F2- CNPAM 93-168	1	0	0	0	0	3	1	5	0	0	0
	5	0	0	0	0	3	1	5	2	0	0
	10	0	0	0	0	3	1	3	2	0	0
	25	0	0	0	0	0	1	0	1	0	0
F3- CNPAM 2012-137	1	0	0	0	0	2	0	0	3	0	0
	5	0	0	0	0	2	0	0	5	0	0
	10	0	0	0	0	1	0	0	5	0	0
	25	0	0	0	0	0	0	0	0	0	0
F3- CNPAM 93-168	1	0	0	0	0	1	2	0	1	0	0
	5	0	0	0	0	1	2	0	1	0	0
	10	0	0	0	0	1	1	0	1	0	0
	25	0	0	0	0	0	0	0	1	0	0
F4- CNPAM 2012-137	1	0	0	0	0	0	0	0	0	0	1
	5	1	0	0	0	0	0	0	0	0	1
	10	0	0	0	0	0	0	0	0	0	1
	25	0	0	0	0	0	0	0	0	1	0
F4- CNPAM 93-168	1	0	0	0	0	0	2	0	3	0	0
	5	0	0	0	0	0	2	0	3	0	0
	10	0	0	0	0	0	2	0	1	0	0
	25	0	0	0	0	0	0	0	1	0	0

Os resultados de erros de classificação são promissores, exceto para o modelo com 25% como nível de significância. Por se tratar de material instável oriundos de gerações recíprocas, os resultados obtidos demonstram a aplicabilidade e funcionalidade do modelo, gerando erros do tipo I menos frequentes. Os erros do Tipo II são menos frequentes nos parentais e maiores para gerações recíprocas frente aos seus genitores.

Neste particular, vale considerar a complexidade da matriz e o número de amostras analisadas. Considerados válido o modelo aplicado e o nível de significância escolhido foi o de 5%, por apresentar as melhores respostas frente aos conjuntos analisados e por ser o mais utilizado na literatura (**Tabela 11**)

Tabela 12. Erros de classificação ao nível de 5% de confiança para teste F.

	AMOSTRAS DE PREVISÃO									
Modelos	CNPAM 2012- 137	CNPAM 93-168	F1- CNPAM 2012- 137	F1- CNPAM 93-168	F2- CNPAM 2012- 137	F2- CNPAM 93-168	F3- CNPAM 2012- 137	F3- CNPAM 93-168	F4- CNPAM 2012- 137	F4- CNPAM 93-168
CNPAM 2012-137	1	0	4	0	0	0	0	0	0	0
CNPAM 93-168	0	0	0	6	0	0	0	0	0	0
F1- CNPAM 2012-137	4	0	0	0	0	0	0	0	0	0
F1- CNPAM 93-168	0	6	0	0	0	0	0	0	0	0
F2- CNPAM 2012-137	0	0	0	0	2	6	0	3	0	0
F2- CNPAM 93-168	0	0	0	0	3	1	5	2	0	0
F3- CNPAM 2012-137	0	0	0	0	2	0	0	5	0	0
F3- CNPAM 93-168	0	0	0	0	1	2	0	1	0	0
F4- CNPAM 2012-137	1	0	0	0	0	0	0	0	0	1
F4- CNPAM 93-168	0	0	0	0	0	2	0	3	0	0

3.3.2. Construção e Validação do Modelo SPA-LDA

3.3.2.1. Validação cruzada

O número ideal de variáveis para o SPA-LDA foi determinado a partir do mínimo da função de custo G , exibido na **figura 53**

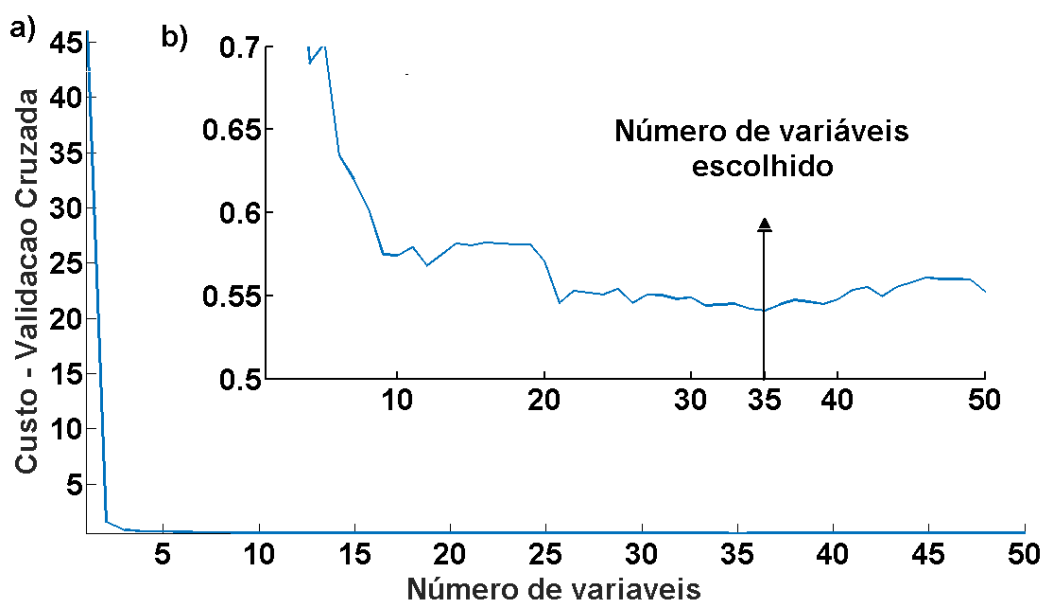


Figura 53- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA por validação cruzada em (a) visualização normal e em (b) com efeito do zoom.

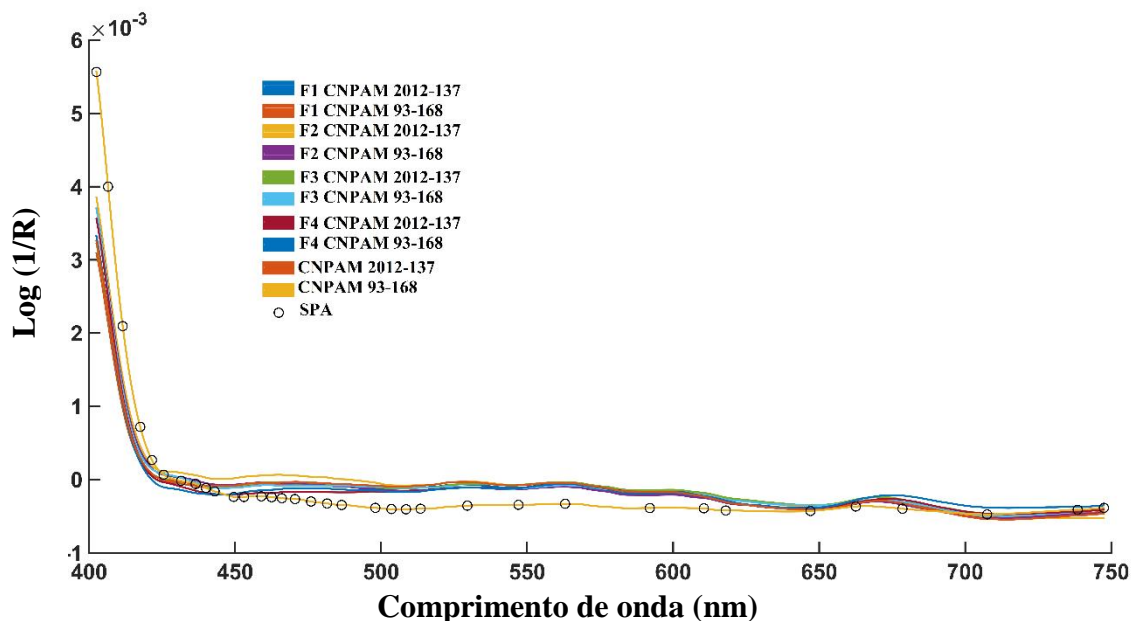


Figura 54- Espectro médio dos genótipos com as variáveis selecionadas pelo SPA-LDA.

De acordo com o gráfico acima (**Figura 54**), é possível observar a aplicação da técnica, destacando as 35 variáveis ao longo dos espectros médios dos genótipos estudados.

Mediante a aplicação do do algoritmo APS, foram selecionadas 35 variáveis como sendo o número ideal, sendo o menor custo obtido no conjunto de teste de 0.54092, apresentando um erro de predição. (**Tabela 13**),

Tabela 13. Resultados do algoritmo SPA-LDA para o conjunto de validação cruzada separado pelo algoritmo KS.

Nome	Referência	Classificação
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	2
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6

F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10

É possível observar na tabela acima em destaque, um erro de classificação, no qual o genótipo F2-CNPAM 93-168 X CNPAM 2012-137, foi classificado como sendo pertencente a F1-CNPAM 93-168 X CNPAM 2012-137.

3.3.2.2. Validação por serie de teste

O número ideal de variáveis para o SPA-LDA foi determinado a partir do mínimo da função de custo G, exibido na **figura 55**.

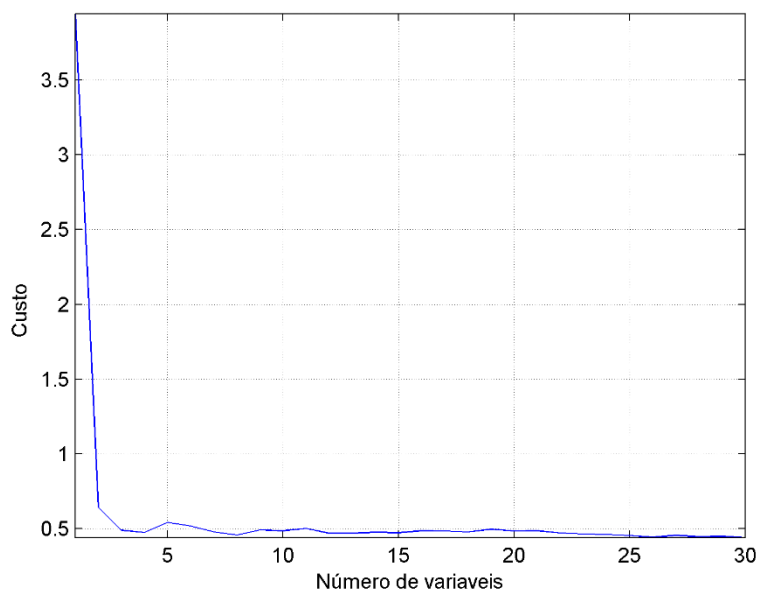


Figura 55- Gráfico da função do custo associado à seleção de variáveis com o SPA-LDA, por série de teste.

Mediante a aplicação do algoritmo APS, foram selecionadas 22 variáveis como sendo o número ideal, o menor custo obtido no conjunto de teste foi de 0.47199, não apresentando erros de teste e nem erros de predição (**Tabela 14**).

Tabela 14. Resultados do algoritmo SPA-LDA para o conjunto de previsão separado pelo algoritmo KS.

Nome	Referência	Classificação
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 2012-137 X CNPAM 93-168	1	1
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F1-CNPAM 93-168 X CNPAM 2012-137	2	2
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3

F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 2012-137 X CNPAM 93-168	3	3
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F2-CNPAM 93-168 X CNPAM 2012-137	4	4
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 2012-137 X CNPAM 93-168	5	5
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F3-CNPAM 93-168 X CNPAM 2012-137	6	6
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 2012-137 X CNPAM 93-168	7	7
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
F4-CNPAM 93-168 X CNPAM 2012-137	8	8
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 2012-137	9	9
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10
CNPAM 93-168	10	10

Na **figura 56** observa-se o espectro médio das amostras de treinamento ao qual foram indexadas as variáveis selecionadas pelo SPA-LDA.

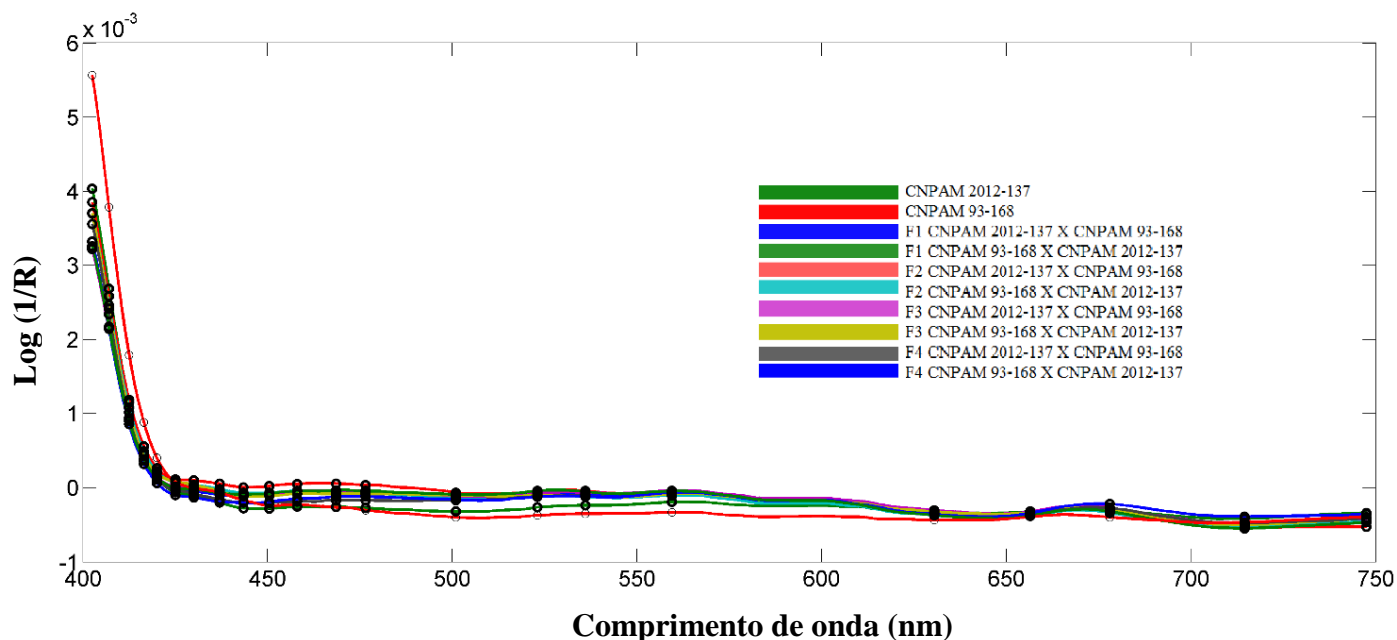


Figura 56- Espectro médio das amostras de treinamento com as variáveis selecionada pelo SPA-LDA.

3.3.3. Aplicação dos Modelos ao Conjunto de Teste

Os parâmetros utilizados nos modelos SIMCA foram: quatro PCs, para cada classe e 5% de nível de significância estatística.

A capacidade de discriminação das etapas de previsão em ambos os modelos foi comprovada no conjunto externo de amostras.

Observamos que o uso do modelo SPA-LDA para esse tipo de matriz, mostrou-se eficaz. Quando feito o uso da validação cruzada foi possível observar um erro de predição, enquanto que, com o uso da série de teste todas as amostras foram classificadas corretamente, sendo este mais recomendado.

4. CONCLUSÕES

- A PCA permitiu discriminar os parentais e as gerações recíprocas de mamona na região espectral correspondente à faixa de 400 a 750nm, sendo registradas áreas de sobreposição em alguns conjuntos de amostras;
- O modelo SIMCA forneceu erros do tipo I e do tipo II sendo em maior número o do tipo II devido à natureza do conjunto amostral, comprovando a eficiência da técnica;
- O SPA-LDA mostrou-se eficiente, selecionando vinte duas variáveis espectrais classificando corretamente todas as amostras do conjunto previsão;
- A técnica permitiu que todas as análises fossem realizadas sem o uso de reagentes, a baixo custo e de forma rápida.
- Os modelos utilizados são promissores na classificação de sementes de mamona na faixa espectral do VIS-NIR, necessitando de maior aprofundamento na determinação de qual ou quais comprimentos de onda são responsáveis pelas melhores respostas;
- Deve-se buscar o desenvolvimento de novos estudos referentes a predição não destrutiva de componentes químicos presentes na amostra e detectáveis na região VIS-NIR para auxiliar na tomada de decisão em programas de melhoramento visando incrementos em gerações subsequentes;
- Maiores estudos devem ser realizados com intuito de identificar a influência do tegumento das sementes nos resultados obtidos;
- Ampliar os estudos com o intuito de correlacionar as gerações as classes certificadas de semente, fornecendo base ao MAPA para análise de pureza em escala.

5. REFERENCIAS BIBLIOGRAFICAS

ALVARENGA, S. M.; CAIXETA, E. T.; OLIVEIRA, A. C. B.; RUFINO, R. J. N.; BRITO, G. G.; PEREIRA, A. A.; ZAMBOLIM, L.; SAKIYAMA, N. S. **Análise da diversidade genética de cafeeiros do banco de germoplasma da UFV/EPAMIG utilizando marcadores RAPD.** In: Simpósio de pesquisas dos cafés do Brasil, 4. Londrina,PR, 2005.

BOUHARMONT, P.; LOTODÉ, R.; AWEMO, J.; CASTAING, X. La sélection generative du caféier robuste au Cameroun. Analyse des resultats d'un essai d'hybrides dialléle partiel implanté en 1973. **Café Cacao Thé**, v.30, n.2, p.93-112, 1986.

CABRAL, T. A. T. **Padrões moleculares, diversidade genética e mapa parcial de ligação do cafeeiro.** Tese (Doutorado) – Universidade Federal de Viçosa, Viçosa, MG. 2001.

COSTA, M.N.; PEREIRA, W. E.; BRUNO, R. L. A.; FREIRE, E. C.; NOBREGA, M. B. M.; MILANI, M.; OLIVEIRA, A. P., Genetic divergence on castor bean accesses and cultivars through multivariate analysis. **Pesquisa Agropecuária Brasileira**, v. 41, 1617-1622, 2006.

CRUZ C. D.; CARNEIRO P. C. S. Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV. 2003.585 p

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético.** Viçosa: UFV, 2004. p.223-375.

CUI, H. F.; YE, Z. H.; XU, L.; FU, X. S.; FAN, C.W.; YU, X. P. Automatic and Rapid Discrimination of Cotton Genotypes by Near Infrared Spectroscopy and Chemometrics. **Journal of Analytical Methods in Chemistry**. Volume 2012, Article ID 793468, 7 pages DOI:10.1155/2012/793468

DESTRO, D.; MONTALVÁN, R. **Melhoramento genético de plantas.** Londrina-PR: UEL. 1999. 749p.

FALCONER, D. S., MACKAY, T. F. **Introduction to quantitative genetics**. 4th ed., Malásia: Longman, 1996. 464p.

FEHR, W. R. (ed.) **Principles of cultivar development**:. theory and technique. Iowa State University Press: Ames, Iowa, 1987. 536p. v.1

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3. ed. Brasília: Embrapa-Cenargen, 1998. 220p.

FIGUEIREDO NETO A. et al. Divergência genética em acessos de mamona (*Ricinus communis* L.) baseada nas características das sementes. **Revista de Biologia e Ciências da Terra**, Campina Grande, v. 4, n. 2, p. 1-10, 2004.

FREIRE, E.C.; LIMA, E.F.; ANDRADE, F.P. Melhoramento genético. In: AZEVEDO, D.M.P. de; LIMA, E.F. (Ed.). O agronegócio da mamona no Brasil. Brasília: Embrapa Informação Tecnológica, 2001. p.229-256.

FUFA, H.; BAENZIGER, P. S.; BEECHER, B. S.; DWEIKAT, I.; GRAYBOSCH, R. A.; ESKRIDGE, K. M. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. **Euphytica**, Wageningen, v. 145, p. 133-146, 2005.

HAN S.I.; CHAE, J. H.; BILYEU, K.; SHANNOM, J. G.; LEE, J. D. Non-destructive Determination of High Oleic Acid Content in Single Soybean Seeds by Near Infrared Reflectance Spectroscopy. **J Am Oil Chem Soc** v. 91, p. 229–234, 2014. DOI 10.1007/s11746-013-2369-y

IVOGLO, M. G. **Divergência genética entre progênies de café robusta**. Campinas-SP. 75fls. Dissertação (Mestrado em Agricultura Tropical e Subtropical) Instituto Agrônômico-IAC, 2007.

LASHERMES, P.; TROUSLOT, F.; COMBES, M. C.; CHARRIE, R. A. Genetic diversity for RAPD markers between cultivated and wild accession of *C. arabica*. **Euphytica**, v.87, n.1 p.59-64, 1996.

MONTAGNON, C., LEROY, T., YAPO, A. Diversité génotypique et phénotypique de quelques groupes de caféiers (*Coffea canéfora* Pierre) en collection. **Café Cacao Thé**, v.36, n.3, p.187-197, 1992.

MOREIRA, J. A. N.; LIMA, E. F.; FARIAS, F. J. C.; AZEVEDO, D. M. P. **Melhoramento da mamoneira (*Ricinus communis* L.)**. Campina Grande: Embrapa Algodão, 29p. Documento, 44. 1996.

MOSHKIN, V. A. **Castor**. New Delhi: Amerind, 1986. 315 p.

NASS, L. L.; VALOIS, C. C.; MELO, I. S.; VALADARES-INGLIS, M. C. (ed.). **Recursos genéticos e melhoramento**. Fundação MT, Rondonópolis-MT: 2001. 1183p.

OLIVEIRA, A. C. B.; SEDIYAMA, M. A. N.; SEDIYAMA, T.; FINGER, F. L.; CRUZ, C. D. Variabilidade genética em batata doce com base em marcadores izoenzimáticos. **Horticultura Brasileira**, v.20, n.4, p.576-582, 2002.

OROZCO-CASTILLO, C.; XHALMERS, K. J.; WAUGH, R.; POWELL, W. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. **Theoretical and Applied Genetics**, v.87, n.8, p.934-940, 1994.

SEVERINO, L. S.; MILANI, M.; MORAES, C. R. de A.; GONDIM, T. M. de S.; CARDOSO, G. D. Avaliação da produtividade e teor de óleo de dez genótipos de mamoneira cultivados em altitude inferior a 300 metros. **Revista Ciência Agronômica**, v.37, n.2, p.188-194, 2006.

SILVESTRINI, M.; JUNQUEIRA, M. G.; FAVARIN, A. C.; GUERREIRO FILHO, O.; MALUF, M. P.; SILVAROLLA, M. B.; COLOMBO, C. **Diversidade genética da espécie *Coffea arábica* avaliada por marcadores microssatélites**. In: Simpósio de pesquisas dos cafés do Brasil, 4. Londrina,PR, 2005.

SUDRÉ, C. P. Divergência genética entre acessos de pimenta e pimentão utilizado técnicas multivariadas. **Horticultura Brasileira**, Brasília, v. 23, n. 1, p. 22-27, jan./mar. 2005.